

روش بوت استرپ بلوک مجزا در آمار فضایی

نصراله ایران پناه و محسن محمدزاده: دانشگاه تربیت مدرس

چکیده

روش بوت استرپ افرون برای برآورد میزان دقت برآوردگرها، هنگام مشاهدات مستقل کاربرد دارد. برای داده های فضایی که بر حسب موقعیت قرار گرفتن آنها در فضای مورد بررسی به یکدیگر وابسته هستند، معمولاً روش بوت استرپ بلوک متحرک مورد استفاده قرار می گیرد. از آنجا که در این روش مشاهدات مرزی نسبت به سایر مشاهدات امکان کمتری برای حضور در بلوکها دارند، در این مقاله روش بوت استرپ بلوک مجزا معرفی و الگوریتمی برای برآورد میزان دقت برآوردگرها ارائه می شود. همچنین کارایی روش بوت استرپ بلوک مجزا با بلوک متحرک در یک بررسی شبیه سازی مورد مقایسه عددی قرار گرفته، نشان داده می شود برآوردگر اریبی میانگین نمونه ای به روش بوت استرپ بلوک مجزا بدون خطا و برآوردگر واریانس آن سازگار است.

مقدمه

در آمار کلاسیک عموماً فرض می شود داده های به دست آمده از جامعه، مستقل اند. در عمل با موارد زیادی مواجه می شویم که داده ها وابسته اند. داده های فضایی [۱] مشاهداتی هستند که وابستگی آنها ناشی از موقعیت شان در فضای بررسی شده است و این وابستگی تابعی از فاصله مشاهدات از یکدیگر است. اغلب استنباط های آمار فضایی مبتنی بر گاوسی بودن میدان تصادفی است، که در عمل ممکن است این شرط برقرار نباشد. در این گونه موارد می توان از الگوریتم بوت استرپ و باز نمونه گیری^۱ از داده های فضایی در استنباط استفاده کرد. افرون [۲] روش بوت استرپ را برای داده های مستقل ارائه کرد؛ در این روش می توان با استفاده از باز نمونه گیری داده ها، اریبی، واریانس و توزیع برآوردگرها را برآورد کرد. این روش برای داده های فضایی به علت وابستگی مشاهدات کاربرد ندارد. هال [۳] دو روش براساس بلوک کردن مشاهدات و موقعیت ها برای حالت خاص داده های موزاییک ارائه کرد، که در آنها همبستگی مشاهدات در موقعیت مجاور مرز بلوکها ساختار واقعی ندارد، به این علت هال [۴] کوشیده است این مشکل را با استفاده از نامساوی بن فرونی به نوعی حل کند. بولمان و کونش [۵] و لاهییری [۶] نیز روش بوت استرپ بلوک متحرک (MBB) را برای داده های فضایی ارائه کردند، که در آن ساختار داده ها در فضای $\mathbb{Z}^d$ مورد توجه قرار گرفته است. در این روش نمونه ای از مشاهدات

واژه های کلیدی: بوت استرپ بلوک مجزا، بوت استرپ بلوک متحرک، α - اختلاط، شبیه سازی مونت کارلو

۱-Resampling

۲-Moving Block Bootstrap

در بلوک‌های متحرک بازنمونه‌گیری می‌شود، به گونه‌ای که هر مشاهده حداقل در یکی از بلوک‌ها قرار گیرد، اما امکان کمتر مشاهدات مرزی در مقایسه با مشاهدات مرکزی برای حضور در بلوک‌ها موجب اریبی برآوردها می‌شود.

در این مقاله برای رفع این مشکل روش بوت‌استرپ بلوک مجزا^۱ (SBB)، برای داده‌های فضایی معرفی می‌شود، که در آن ابتدا موقعیت‌ها به بلوک‌های یک‌سان افراز می‌شوند، سپس الگوریتم بوت‌استرپ با استفاده از بازنمونه‌گیری بلوک‌های مجزا اجرا می‌شود. برای این منظور، الگوریتم بوت‌استرپ IID در بخش ۲ ارائه می‌شود. در بخش ۳ روش بوت‌استرپ بلوک مجزا ارائه و روش بوت‌استرپ بلوک متحرک معرفی می‌شود. در بخش ۴ خواص برآوردهای اریبی و واریانس میانگین نمونه‌ای به روش بوت‌استرپ بلوک مجزا مورد بررسی قرار می‌گیرد. نتایج حاصل از دو روش بلوک مجزا و متحرک با استفاده از تکنیک شبیه‌سازی در بخش ۵ مورد مقایسه عددی قرار گرفته و در بخش نهایی به بحث و نتیجه‌گیری پرداخته شده است.

بوت استرپ IID

فرض کنید $Z \equiv \{Z_1, \dots, Z_n\}$ مجموعه‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع تجمعی F_θ و کمیت تصادفی مورد نظر $T = t(Z; F_\theta)$ باشد. روش بوت‌استرپ بر اساس ایده بازنمونه‌گیری از داده‌ها برای تعیین مشخصات توزیع نمونه‌ای T بدون فرض معلوم بودن F_θ است. افرون [۲] الگوریتمی برای برآورد مشخصات توزیع T به عنوان برآوردها پارامتر مورد نظر θ و بر اساس مشاهدات مستقل به صورت زیر پیشنهاد کرد:

(الف) تابع توزیع تجربی $F_n(z) = \frac{1}{n} \sum_{i=1}^n I(Z_i \leq z)$ را تعیین کنید.

(ب) $Z^* \equiv (Z_1^*, \dots, Z_n^*)$ را به عنوان نمونه بوت‌استرپ از F_n به روش نمونه‌گیری تصادفی ساده باجای‌گذاری از Z به دست آورید.

(ج) آماره بوت‌استرپ $T^* = t(Z^*; F_n)$ را محاسبه کنید.

(د) بوت‌استرپ اریبی، واریانس و توزیع T به ترتیب به صورت:

$$G_*(t) = P(T^* \leq t) \text{ و } Var_*(T^*) = E_*[T^* - E_*(T^*)]^2, Bias_*(T^*) = E_*(T^*) - T$$

برآورد می‌شوند، که در آن E_* ، Var_* و P_* به ترتیب امید ریاضی، واریانس و احتمال شرطی بوت‌استرپ به شرط Z است. اگر $Bias_*(T^*)$ ، $Var_*(T^*)$ و $G_*(t)$ جواب‌های صریحی نداشته باشند، با

استفاده

از

1-Separate Block Bootstrap

شبیه سازی مونت کارلو و تکرار B بار مراحل ب و ج و محاسبه $T_1^*, \dots, T_B^*$ به ترتیب به صورت:

$$\widehat{Bias}^*(T^*) = \hat{E}_*(T^*) - T$$

$$\widehat{Var}^*(T^*) = \frac{1}{B-1} \sum_{i=1}^B [T_i^* - \hat{E}_*(T^*)]^2$$

$$\hat{G}_*(t) = \frac{1}{B} \sum_{i=1}^B I(T_i^* \leq t)$$

برآورد شوند، که در آن ها $\hat{E}_*(T^*) = \frac{1}{B} \sum_{i=1}^B T_i^*$ است.

در اغلب تحلیل های نظری بوت استرپ برای F_θ و T شرایطی منظور می شود، به طوری که با افزایش n توزیع و گشتاورهای T^* به توزیع و گشتاورهای مجانبی T میل کند.

بوت استرپ فضایی

معمولاً میدان تصادفی $\{Z(s): s \in D\}$ به عنوان مدل آماری برای تجزیه و تحلیل داده های فضایی در نظر گرفته می شود، که در آن D يك مجموعه اندیس گذار در فضای اقلیدسی $d \geq 1$ بعدی $\mathbb{R}^d$ است. فرض کنید ناحیه نمونه گیری D_n وقتی که $n \rightarrow \infty$ غیر کران دار باشد. گیریم $\tilde{D} \subset \left[-\frac{1}{2}, \frac{1}{2}\right]^d$ يك مجموعه باز هم بند شامل مبدأ و D_o يك مجموعه نمونه اولیه برای نواحی نمونه گیری با شرط $\tilde{D} \subset D_o \subset cl.(\tilde{D})$ باشد، که در آن $cl.(\tilde{D})$ بستار مجموعه $\tilde{D}$ است. همچنین با فرض آن که $\{\lambda_n\}_{n \geq 1}$ يك دنباله صعودی از اعداد حقیقی بزرگ تر از يك باشد، ناحیه نمونه گیری D_n با تکثیر مجموعه نمونه اولیه D_o و ثابت مقیاس بندی λ_n به صورت $D_n = \lambda_n D_o$ در نظر گرفته می شود. فرض کنید در ناحیه نمونه گیری D_n ، تعداد N_n موقعیت محدود $S_n \equiv \{s_1, \dots, s_{N_n}\}$ از يك میدان تصادفی مانای $\{Z(s): s \in \mathbb{Z}^d\}$ به صورت يك شبکه منظم مشاهده شده است. اندازه نمونه N_n و حجم ناحیه نمونه گیری D_n به صورت:

$$N_n = Vol.(D_o) \cdot \lambda_n^d \quad (1)$$

با هم در ارتباط هستند. با قرار دادن $N = N_n$ ، فرض کنید $T_n = t_n(Z_n)$ برآوردگر پارامتر θ باشد، که در آن $Z_n = (Z(s_1), \dots, Z(s_N))$ يك نمونه از مشاهدات است. هدف تعیین مشخصات توزیع نمونه ای T_n به

روش بوت‌استرپ بلوک فضایی است. در این بخش روش جدید بوت‌استرپ بلوک مجزا معرفی شده و با استفاده از نمادهای آن روش بلوک متحرك نیز بیان می‌شود.

۱- بوت‌استرپ بلوک مجزا

فرض کنید $\{\beta_n\}_{n \geq 1}$ يك دنباله از اعداد صحیح مثبت باشد، به طوری که:

$$\beta_n^{-1} + \lambda_n^{-1} \beta_n = o(1) \quad (۲)$$

یعنی β_n با نرخ آهسته‌تری از عامل مقیاس‌بندی λ_n برای ناحیه نمونه‌گیری D_n به بی‌نهایت میل کند. عامل مقیاس‌بندی برای بلوک‌ها یا زیر نواحی در روش بوت‌استرپ بلوک فضایی است. ابتدا ناحیه نمونه‌گیری D_n را با استفاده از مکعب‌های به حجم β_n^d افراز می‌کنیم. گیریم $K_n = \{k \in \mathbb{Z}^d : \beta_n(k+U) \cap D_n \neq \emptyset\}$ مجموعه اندیس مکعب‌های به شکل $\beta_n(k+U)$ باشد که اشتراك غیر تهی با ناحیه نمونه‌گیری D_n دارد و در آن $U = [0,1]^d$ مکعبی واحد در $\mathbb{R}^d$ است. با فرض $N = \beta_n^d |K_n|$ ، ناحیه نمونه‌گیری D_n با مکعب‌های به حجم β_n^d و تعداد $|K_n|$ کامل می‌شود. سپس برای هر $k \in K_n$ ، نمونه بوت‌استرپ را در زیر ناحیه k ام به صورت:

$$D_n(k) \equiv \beta_n(k+U) \cap D_n \quad (۳)$$

در نظر می‌گیریم $I_n = \{i \in \mathbb{Z}^d : \beta_n(i+U) \subset D_n\}$ مجموعه اندیس مکعب‌های مجزا به حجم β_n^d در D_n با نقاط شروع $i \in \mathbb{Z}^d$ باشد، در نتیجه $\{\beta_n(i+U) : i \in I_n\}$ يك مجموعه از زیر نواحی یا بلوک‌های مکعبی مجزا واقع در D_n هستند. فرض کنید $Z_n(D_n) = \{Z(s_1), \dots, Z(s_N)\}$ نمونه کامل و $Z_n(D_n(k))$ زیر نمونه قرار گرفته در زیر ناحیه $D_n(k)$ باشد. برای به دست آوردن يك نمونه بوت‌استرپ بلوک مجزای فضایی، ابتدا برای هر $k \in K_n$ يك بلوک به صورت تصادفی از مجموعه $\{\beta_n(i+U) : i \in I_n\}$ و مستقل از بلوک‌های دیگر انتخاب می‌کنیم. سپس با استفاده از مشاهدات در زیر ناحیه باز نمونه‌گیری شده، نمونه بوت‌استرپ را روی زیر ناحیه $D_n(k)$ تعیین می‌کنیم. سرانجام، فرض کنید $K = K_n$ اندازه مجموعه K_n و همچنین $\{I_k : k \in K_n\}$ يك مجموعه از K متغیر iid با توزیع مشترك

$$P(I_1 = i) = \frac{1}{|I_n|}, \quad i \in I_n$$

باشد. برای هر $k \in K_n$ زیر نمونه بوت‌استرپ بلوک مجزای $Z_n^*(D_n(k))$ را با استفاده از بلوک باز نمونه‌گیری شده $Z_n(\beta_n[I_k + U])$ به صورت:

$$Z_n^*(D_n(k)) = Z_n(\beta_n[I_k + U] \cap [D_n(k) - k\beta_n + I_k\beta_n])$$

به دست می آوریم. حال نمونه بوت استرپ بلوک مجزا $Z_n^*(D_n)$ را از به هم پیوستن بلوک های باز نمونه گیری شده $\{Z_n^*(D_n(k)) : k \in K_n\}$ تعیین و نسخه بوت استرپ بلوک مجزای آماره $T_n = t_n(Z_n)$ را به صورت $T_n^* = t_n(Z_n^*(D_n))$ تعریف می کنیم. در واقع نمونه بوت استرپ بلوک مجزای $Z_n^*(D_n)$ يك نمونه تصادفی ساده با جای گذاری به حجم K از بلوک های مکعبی مجزای $Z_n^*(D_n(k))$ به حجم β_n^d است. ادامه روش مانند مرحله (د) بوت استرپ IID است.

۲- بوت استرپ بلوک متحرك

بولمان و کونش [۵] و لاهیری [۶] روش بوت استرپ بلوک متحرك یا متداخل را برای برآورد مشخصات توزیع T_n ارائه کردند. فرض کنید زیر نواحی $D_n(k)$ افراز ناحیه نمونه گیری D_n داده شده در (۳) به صورت مکعب های به حجم β_n^d باشد. در این روش ابتدا با استفاده از مجموعه مکعب های $J_n = \{j \in \mathbb{Z}^d : (j + \beta_n U) \subset D_n\}$ که متداخل هستند، K متغیر تصادفی iid $\{J_k : k \in K_n\}$ با توزیع مشترک $P(J_1 = j) = \frac{1}{|J_n|}$ ، $j \in J_n$ تولید می شود. سپس زیر نمونه بوت استرپ بلوک متحرك به صورت:

$$Z_n^{**}(D_n(k)) = Z_n([J_k + \beta_n U] \cap [D_n(k) - k\beta_n + J_k])$$

به دست می آید. حال نسخه بوت استرپ بلوک متحرك آماره T_n به صورت $T_n^{**} = t_n(Z_n^{**}(D_n))$ تعریف می شود، که در آن $Z_n^{**}(D_n)$ از به هم پیوستن بلوک های باز نمونه گیری شده $\{Z_n^{**}(D_n(k)) : k \in K_n\}$ تعیین می شوند. در واقع نمونه بوت استرپ بلوک متحرك $Z_n^{**}(D_n)$ يك نمونه تصادفی ساده با جای گذاری به حجم K از بلوک های مکعبی متداخل $Z_n^{**}(D_n(k))$ به حجم β_n^d است.

خواص برآوردگرهای بوت استرپ اریبی و واریانس میانگین نمونه ای

فرض کنید برای میدان تصادفی مانای $\{Z(i) : i \in \mathbb{Z}^d\}$ ، میانگین نمونه ای $\bar{Z}_n = N^{-1} \sum_{i=1}^N Z(s_i)$ به عنوان برآوردگر میانگین میدان تصادفی $[Z(0)]$ ، $\mu = E[Z(0)]$ ، براساس مشاهدات Z_n باشد. اگر $\bar{Z}_n^*$ میانگین نمونه ای بوت استرپ بلوک مجزا باشد، در این صورت برآورد بوت استرپ پارامتر $Bias(\bar{Z}_n) = E(\bar{Z}_n) - \mu = 0$ به صورت $Bias(\bar{Z}_n^*) = E(\bar{Z}_n^*) - \bar{Z}_n$ تعیین می شود. **لم ۱:** روش بوت استرپ بلوک مجزا، اریبی میانگین نمونه ای از يك میدان تصادفی مانای $\{Z(i) : i \in \mathbb{Z}^d\}$ را بدون خطا برآورد می کند.

اثبات: مجموعه‌های $S_n(i; k) = \sum_{j \in B_n(i; k)} Z(j)$ و $S_n^*(k) = \sum_{j \in B_n(I_k; k)} Z^*(j)$ که در آن‌ها $B_n(i; k) \equiv D_n(k) - k\beta_n + i\beta_n$ است $(k \in K_n, i \in I_n)$ را در نظر بگیرید. به دلیل شانس حضور یکسان مشاهدات در بلوک‌های مجزای $\{\beta_n(i+U): i \in I_n\}$ و با توجه به $|K_n| = |I_n|$ داریم:

$$\begin{aligned} E_*(\bar{Z}_n^*) &= E_* \left[N^{-1} \sum_{k \in K_n} S_n^*(k) \right] \\ &= N^{-1} |K_n| E_*[S_n^*(\cdot)] \\ &= N^{-1} |K_n| |I_n|^{-1} \sum_{i \in I_n} S_n(i; \cdot) \\ &= N^{-1} \sum_{i=1}^N Z(s_i) = \bar{Z}_n \end{aligned}$$

بنا بر این، در روش SBB ارببی میانگین نمونه‌ای بدون خطا برآورد می‌شود.

در روش MBB برخلاف روش SBB به دلیل امکان نابرابر مشاهدات برای حضور در بلوک‌های متحرک $\{(j + \beta_n U): j \in J_n\}$ برآورد ارببی میانگین نمونه‌ای همراه با خطاست. برای رفع این مشکل معمولاً با مرکزی کردن آماره به شکل $T_n = \bar{Z}_n - \mu$ برآورد بوت استرپ آن به صورت $T_n^* = \bar{Z}_n^* - E_*(\bar{Z}_n^*)$ در نظر گرفته می‌شود.

برآورد بوت استرپ پارامتر $\sigma_n^2 = N \text{Var}(\bar{Z}_n)$ را می‌توان بر اساس پارامتر اندازه-بلوک β_n به صورت

$$\hat{\sigma}_n^2 = \hat{\sigma}_n^2(\beta_n) = N \text{Var}_*(\bar{Z}_n^*) \quad (۴)$$

تعیین کرد. برای بررسی سازگاری $\hat{\sigma}_n^2$ ، لازم است ابتدا اندازه وابستگی برای میدان تصادفی تعریف شود. واضح است برای هر دو زیر- σ -جبر مستقل A و B از F در یک فضای احتمال (Ω, F, P) برای هر $A \in \mathcal{A}$ و $B \in \mathcal{B}$ ، $P(A \cap B) - P(A)P(B) = 0$ است. اما وقتی A و B مستقل نباشند، معمولاً اندازه وابستگی A و B با بزرگترین مقدار وابستگی Δ به صورت:

$$\alpha(A, B) = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{A}, B \in \mathcal{B} \}$$

تحت عنوان ضریب وابستگی α -اختلاط (α -mixing) تعیین می‌شود. برای میدان تصادفی $\{Z(s): s \in \mathbb{R}^d\}$ به ازای $a \geq 0$ و $b \geq 1$ ضریب α -اختلاط را به صورت:

$$\alpha(a; b) = \sup \{ \alpha_1(S_1, S_2) : S_1, S_2 \in \mathcal{R}(b), d(S_1, S_2) \geq a \}$$

تعریف می‌کنیم، که در آن $\mathcal{R}(b)$ گردایه تمام زیر مجموعه‌های $\mathbb{R}^d$ با حجم b یا کمتر،

$$d(S_1, S_2) = \inf \left\{ \sum_{i=1}^d |x_i - y_i| : (x_1, \dots, x_d)' \in S_1, (y_1, \dots, y_d)' \in S_2 \right\}$$

فاصله مجموعه های S_1 و S_2 است و

$$\alpha_1(S_1, S_2) = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_Z(S_1), B \in \mathcal{F}_Z(S_2) \}$$

است، به طوری که $F_Z(S)$ ، σ -جبر تولید شده توسط متغیر تصادفی $\{Z(s) : s \in S \cap \mathbb{Z}^d\}$ است.

لم ۲: میدان تصادفی $\{Y(i) : i \in \mathbb{Z}^d\}$ را با فرض $E[Y(i)] = 0$ و $E|Y(i)|^{2q+\delta} < \infty$

برای هر $i \in \mathbb{Z}^d$ و عددی طبیعی مانند q و $\delta \in (0, \infty)$ در نظر بگیرید. فرض کنید

$$\sum_{r=0}^{\infty} (r+1)^{d(2q-1)-1} [\alpha_Y(r; 2q)]^{\delta/(2q+\delta)} < \infty$$

که در آن $\alpha_Y(\cdot, \cdot)$ ضریب α -اختلاط $Y(\cdot)$ است. آنگاه برای هر زیر مجموعه $A \subset \mathbb{Z}^d$

$$E \left| \sum_{i \in A} Y(i) \right|^{2q} \leq C(q, \delta) \max \left\{ \left[\sum_{i \in A} \left(E|Y(i)|^{2+\delta} \right)^{\frac{2}{2+\delta}} \right]^q, \sum_{i \in A} \left(E|Y(i)|^{2q+\delta} \right)^{\frac{2q}{2q+\delta}} \right\}$$

که در آن ثابت $C(q, \delta)$ تنها به q ، δ و d و ضریب α -اختلاط $\alpha_Y(\cdot, \cdot)$ بستگی دارد، ولی به زیر

مجموعه A بستگی ندارد.

اثبات: برای اثبات به قضیه ۱.۴.۱.۱ دخان [۷] مراجعه شود.

همان طور که لاهیری [۶] سازگاری برآوردگر واریانس میانگین نمونه ای به روش بوت استرپ بلوک

متحرک را نشان داد، این خاصیت برای روش بوت استرپ بلوک مجزا نیز در قضیه زیر اثبات می شود.

قضیه ۱: فرض کنید میدان تصادفی $\{Z(i) : i \in \mathbb{Z}^d\}$ مانا با $E|Z(\cdot)|^{6+\delta} < \infty$ و ضریب α -اختلاط

$\alpha(a; b) \leq ca^{-\tau_1} b^{\tau_2}$ برای $a \geq 1$ و $b \geq 1$ و به ازای عددی نامنفی مانند δ ، $\tau_1 \geq \frac{5d(6+\delta)}{\delta}$ و

$0 \leq \tau_2 \leq \frac{\tau_1}{d}$ باشد. اگر $\beta_n^{-1} + \lambda_n^{-1} \beta_n = o(1)$ آنگاه

$$\hat{\sigma}_n^2 \xrightarrow{P} \sigma_\infty^2, \text{ as } n \rightarrow \infty$$

که در آن $\hat{\sigma}_n^2$ در (۴) تعریف شده و

$$\begin{aligned} \sigma_\infty^2 &= \lim_{n \rightarrow \infty} N \text{Var}(\bar{Z}_n) \\ &= \sum_{i \in \mathbb{Z}^d} E[Z(0) - \mu][Z(i) - \mu]. \end{aligned}$$

اثبات: بدون از دست دادن کلیت مسئله فرض می شود $\mu = 0$. با توجه به استقلال بلوک های باز نمونه گیری

شده و $N = \beta_n^d |K_n|$ داریم

$$\begin{aligned}
N\text{Var}_*(\bar{Z}_n^*) &= N\text{Var}_*\left(N^{-1} \sum_{k \in K_n} S_n^*(k)\right) \\
&= N^{-1} \sum_{k \in K_n} \text{Var}_*[S_n^*(k)] \\
&= \beta_n^{-d} \left[E_*(S_n^*(\cdot))^2 - (E_*(S_n^*(\cdot)))^2 \right] \\
&\equiv \hat{Q}_{1n} - \hat{Q}_{2n}
\end{aligned} \tag{5}$$

همچنین داریم

$$\begin{aligned}
E(\hat{Q}_{1n}) &= E\left[\beta_n^{-d} E_*(S_n^*(\cdot))^2\right] \\
&= \beta_n^{-d} E[S_n(i; \cdot)]^2 \\
&= \beta_n^{-d} E\left[\sum_{j \in \beta_n U \cap \mathbb{Z}^d} Z(j)\right]^2 \\
&= \sum_{j \in \beta_n U \cap \mathbb{Z}^d} E[Z(0)Z(j)] \rightarrow \sigma_\infty^2 \quad \text{as } n \rightarrow \infty
\end{aligned} \tag{6}$$

با توجه به لم ۱ داریم

$$\begin{aligned}
E(\hat{Q}_{2n}) &= E\left[\beta_n^{-d} (E_*(S_n^*(\cdot)))^2\right] \\
&= \beta_n^{-d} E\left[N |K_n|^{-1} \bar{Z}_n\right]^2 \\
&= \beta_n^d \text{Var}[\hat{Z}_n(s_0)] \\
&= O(\lambda_n^{-d} \beta_n^d)
\end{aligned} \tag{7}$$

بنا بر روابط (۵) تا (۷) برآوردگر $\hat{\sigma}_n^2$ ناریب است. با استفاده از لم ۲ برای $\delta = q = 1$ داریم

$$\begin{aligned}
\text{Var}(\hat{Q}_{1n}) &= \text{Var}\left[\beta_n^{-d} E_*\left(S_n^*(\cdot)\right)^2\right] \\
&= \beta_n^{-2d} \text{Var}\left[\left|I_n\right|^{-1} \sum_{i \in I_n} \left(S_n(i, \cdot)\right)^2\right] \\
&= N^{-2} E\left[\sum_{i \in I_n} \left(S_n(i, \cdot)\right)^2 - E\left(S_n(i, \cdot)\right)^2\right]^2 \\
&= N^{-2} E\left[\sum_{i \in I_n} V_n(i)\right]^2 \\
&\leq N^{-2} |I_n| \max\left\{\left(E\left|V_n(i)\right|^3\right)^{2/3} : i \in I_n\right\} \sum_{r=0}^{\infty} (r+1)^{d-1} \alpha(r \beta_n; \beta_n^d)^{1/3} \\
&\leq N^{-2} |I_n| \max\left\{E\left|S_n(i; 0)\right|^6 : i \in I_n\right\}^{2/3} \sum_{r=0}^{\infty} (r+1)^{d-1} (r+1)^{-\tau_1/3} \left(\beta_n^{\tau_2 d - \tau_1}\right)^{1/3} \\
&= O\left(N^{-2} |I_n| \left(\beta_n^{3d}\right)^{2/3}\right) \\
&= O\left(\lambda_n^{-d} \beta_n^d\right)
\end{aligned} \tag{۸}$$

• ب

طور مشابه داریم

$$\begin{aligned}
\text{Var}(\hat{Q}_{2n}) &= \text{Var}\left[\beta_n^{-d} \left(E_*\left(S_n^*(\cdot)\right)\right)^2\right] \\
&= \beta_n^{-2d} \text{Var}\left[\beta_n^{2d} \hat{Z}_n(s_0)\right]^2 \\
&= \beta_n^{2d} \text{Var}\left[\hat{Z}_n^2(s_0)\right] \\
&= O\left(\lambda_n^{-2d} \beta_n^{2d}\right)
\end{aligned} \tag{۹}$$

بنا بر روابط (۸) و (۹) وقتی $n \rightarrow \infty$ ، آنگاه $\text{Var}(\hat{Q}_{1n}) \rightarrow 0$ و $\text{Var}(\hat{Q}_{2n}) \rightarrow 0$. بنا بر این با توجه به ناریبی مجانبی، $\hat{\sigma}_n^2$ برآوردگر سازگار σ_∞^2 است.

لازم به ذکر است که نحوه تحقق شرایط قضیه ۱ در عمل، در مورد ضریب α -اختلاط در دخان [۷] مورد بررسی قرار گرفته است.

مقایسه بوت استرپ بلوک مجزا و متحرك

در این بخش، دو روش SBB و MBB با استفاده از تکنیک شبیه سازی مونت کارلو برای داده های فضایی (SMS) مورد مقایسه عددی قرار می گیرند. فرض کنید $\{Z(s) : s \in \mathbb{Z}^2\}$ یک میدان تصادفی گاوسی مانای مرتبه دوم با میانگین صفر و هم تغییرنگار نمایی

$$\sigma(h; \theta) = \begin{cases} c_0 + c & h = 0 \\ ce^{-\frac{h}{a}} & h \neq 0 \end{cases}$$

باشد. با فرض $\theta = (c_0, c, a) = (1, 1, 1)$ و داده‌های فضایی در یک شبکه منظم مستطیلی 20×30 ($N = 600$) قرار داشته باشند، با در نظر گرفتن فاصله اقلیدسی بین دو موقعیت، ماتریس کواریانس میدان تصادفی به صورت $\Sigma = (\sigma(s_i - s_j))$ به دست می‌آید. اگر $\Sigma^{1/2}$ تجزیه چولسکی^۱ ماتریس Σ به صورت $\Sigma = \Sigma^{1/2} \times \Sigma^{1/2}$ باشد، یک نمونه تصادفی $Z_n = (Z(s_1), \dots, Z(s_N))$ از میدان تصادفی را می‌توان به صورت $Z_n = \Sigma^{1/2} X_n$ به دست آورد. که در آن $X_n = (X_1, \dots, X_N)$ یک نمونه تصادفی (iid) از توزیع $N(0, 1)$ است. اگر $T_n = t_n(Z_n)$ برآوردگر پارامتر θ باشد، با تکرار B بار (به قدر کافی بزرگ) نمونه‌گیری به حجم N و محاسبه مقادیر $T_1, \dots, T_B$ می‌توان اریبی، واریانس و توزیع

$$T_n = \sqrt{N} \bar{Z}_n$$

$$\text{Bias}(T_n) \approx \bar{T} - \theta$$

$$\text{Var}(T_n) \approx \frac{1}{B-1} \sum_{i=1}^B (T_i - \bar{T})^2$$

$$G(t) \approx \frac{1}{B} \sum_{i=1}^B I(T_i \leq t)$$

تقریب کرد، که در آن‌ها $\bar{T} = \frac{1}{B} \sum_{i=1}^B T_i$ است. مقادیر واقعی اریبی و واریانس آماره $T_n = \sqrt{N} \bar{Z}_n$ که در آن $\bar{Z}_n$ میانگین نمونه‌ای به عنوان برآوردگر پارامتر $\theta = \sqrt{N} \mu$ است، برای 10000 بار تکرار نمونه‌گیری به حجم $N = 600$ محاسبه و در ستون SMS جدول ۱ نشان داده شده است. مقادیر برآورد اریبی

و

۱-Choleski Decomposition

واریانس آماره T_n به دو روش بوت استرپ بلوک مجزا و متحرک به ترتیب در دو ستون SBB و MBB جدول ۱ نشان داده شده است. برای انجام روش بوت استرپ بلوک فضایی، ناحیه نمونه‌گیری را به صورت $D_n = [-10, 10] \times [-15, 15]$ با ثابت مقیاس‌بندی $\lambda_n = 30$ ، اندازه بلوک $\beta_n = 5$ و مجموعه نمونه اولیه $D_0 = \left[-\frac{1}{3}, \frac{1}{3}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right] \subset \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ در نظر می‌گیریم. در این دو روش که با استفاده از مجموعه مشاهده

شده $z_n = \{z(s_1), \dots, z(s_N)\}$ به حجم $N = 600$ که در يك شبکه منظم مستطیلی 20×30 از میدان تصادفی مورد نظر قرار گرفته‌اند، بلوک‌هایی به ابعاد 5×5 را به دو صورت مجزا و متحرك در نظر می‌گیریم. برای این منظور ناحیه نمونه‌گیری را به 24 زیرناحیه یا بلوک به صورت

$$D_n(k) = [5k_1, 5k_1 + 5) \times [5k_2, 5k_2 + 5), k = (k_1, k_2)' \in \mathbb{Z}^2, -2 \leq k_1 < 2, -3 \leq k_2 < 3$$

افراز می‌کنیم. در روش SBB، بازنمونه‌گیری از بلوک‌های مجزای

$$\{[5i_1, 5i_1 + 5) \times [5i_2, 5i_2 + 5), i = (i_1, i_2)' \in \mathbb{Z}^2, -2 \leq i_1 < 2, -3 \leq i_2 < 3\}$$

و در روش MBB، بازنمونه‌گیری از بلوک‌های متحرك

$$\{[j_1, j_1 + 5) \times [j_2, j_2 + 5), j = (j_1, j_2)' \in \mathbb{Z}^2, -10 \leq j_1 < 5, -15 \leq j_2 < 10\}$$

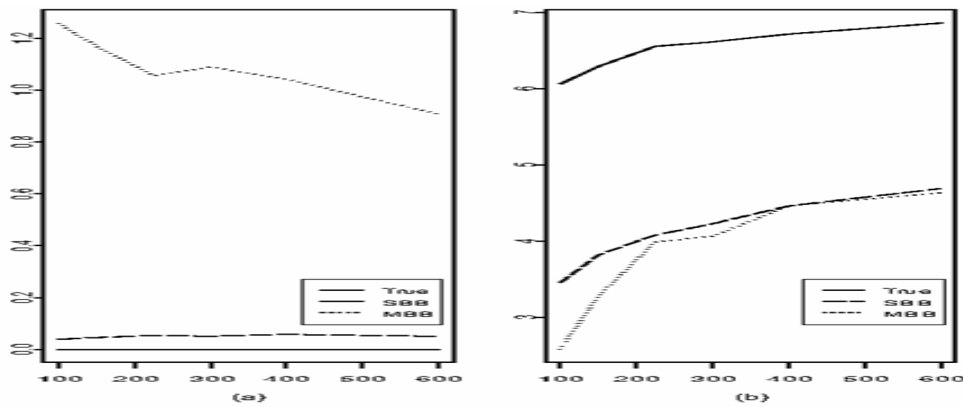
انجام می‌شود. تعداد کل بلوک‌های مجزا و متحرك در این شبکه به ترتیب 24 و 416 است که از بین آن‌ها به روش نمونه‌گیری تصادفی ساده با جایگذاری، 24 بلوک به حجم 25 را انتخاب و از به هم پیوستن آن‌ها يك نمونه بوت‌استرپ $Z_n^* = \{Z^*(s_1), \dots, Z^*(s_N)\}$ به حجم $N = 600$ تولید می‌شود. با محاسبه آماره بوت‌استرپ $T_n^* = \sqrt{N} \bar{Z}_n^*$ و تکرار 2000 بار این الگوریتم، $T_1^*, \dots, T_{2000}^*$ به دست می‌آید. اریبی و واریانس آماره T_n با استفاده از شبیه‌سازی مونت کارلو به روش بوت‌استرپ برآورد شد.

جدول ۱- مقادیر واقعی اریبی و واریانس آماره و برآورد آن‌ها

اندازه دقت	SMS	SBB	MBB
اریبی	۰/۰۰۰	۰/۰۵۱	۰/۹۰۷
واریانس	۶/۸۶	۴/۷۸	۴/۶۷

همان‌طور که در جدول ۱ ملاحظه می‌شود، روش SBB نسبت به روش MBB مقدار اریبی آماره T_n را بسیار نزدیک به مقدار واقعی SMS برآورد می‌کند. در حالی‌که مقدار واریانس برآورد شده آماره T_n به هر دو روش SBB و MBB تقریباً یکسان و نزدیک به مقدار واقعی SMS است. همچنین برای بررسی خاصیت سازگاری و تأثیر اندازه نمونه N در برآوردهای اریبی و واریانس آماره T_n به دو روش SBB و MBB، مقادیر جدول ۱ را به طور مشابه برای اندازه‌های نمونه 100، 150، 225، 300، 400 و 600 در شبکه‌هایی منظم به ترتیب با ابعاد 10×10 ، 10×15 ، 15×15 ، 15×20 ، 20×20 و 20×30 و بلوک‌هایی به ابعاد 5×5 محاسبه کرده‌ایم. شکل ۱ مقدار واقعی و برآوردهای اریبی و واریانس میانگین نمونه‌ای دو روش SBB و MBB را به ازای مقادیر مختلف N ، به ترتیب به صورت خط ممتد، خط چین و نقطه چین نمایش می‌دهد. همان‌طور که در شکل (a) ۱ ملاحظه می‌شود، برآورد اریبی به روش SBB تقریباً

بدون خطا، نزدیک مقدار واقعی اریبی است. در حالی که برآورد اریبی به روش MBB همراه با خطاست، ولی با افزایش اندازه نمونه، به مقدار واقعی صفر نزدیک می‌شود. همچنین در شکل ۱(b) ملاحظه می‌شود که برآورد واریانس به روش SBB نسبت به MBB تقریباً نزدیکتر به مقدار واقعی واریانس است و هر دو برآورد با افزایش اندازه نمونه به مقدار واقعی واریانس نزدیک می‌شوند.



شکل ۱- (a) مقادیر واقعی اریبی و برآورد آن‌ها. (b) مقادیر واقعی واریانس و برآورد آن‌ها

چون مقادیر این برآوردها به نمونه مشاهده شده حساس هستند، یعنی با تعویض نمونه مقدار آن‌ها تغییر خواهد کرد، برای مقایسه اندازه‌های دقت دو روش SBB و MBB از ملاک میانگین مربعات خطا (MSE) حاصل از 1000 بار تکرار برآورد بوت‌استرپ اریبی و واریانس آماره T_n به صورت

$$MSE = \frac{1}{1000} \sum_{i=1}^{1000} [\hat{R}_*(i) - R]^2$$

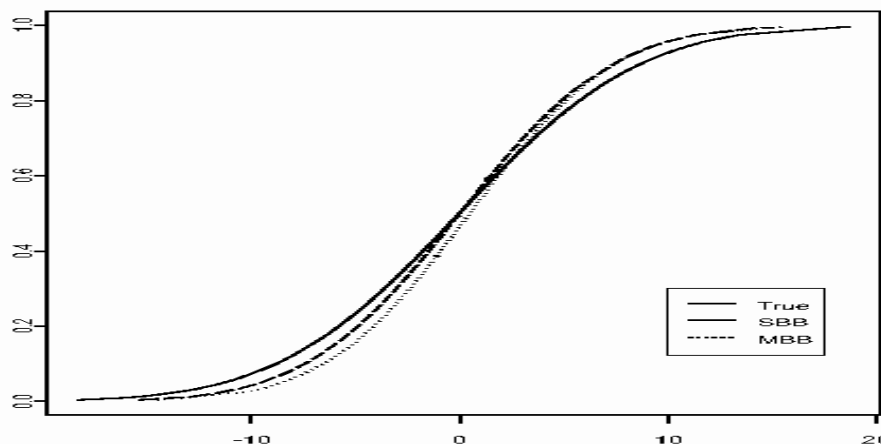
استفاده شده است، که در آن R مقدار واقعی اندازه دقت آماره T_n است و $\hat{R}_*(i)$ نشان دهنده برآورد بوت‌استرپ این اندازه‌ها در تکرار i ام است. با مقایسه مقادیر MSE برای دو روش SBB و MBB که در جدول ۲ نشان داده شده‌اند، ملاحظه می‌شود که برآورد اریبی به روش SBB نسبت به MBB از دقت بسیار بیش‌تری برخوردار است، در حالی که دقت برآورد واریانس به هر دو روش SBB و MBB تقریباً یکسان است.

جدول ۲- مقادیر MSE برآوردهای اریبی و واریانس

MBB	SBB	اندازه دقت
۱/۲۷۳	۰/۰۰۴	اریبی

واریانس	۶/۵۱	۶/۹۵
---------	------	------

همچنین مقدار واقعی توزیع آماره T_n و برآورد آن به دو روش SBB و MBB در شکل ۲ نشان داده شده است. همان‌طور که ملاحظه می‌شود، برآورد توزیع به روش SBB نسبت به روش MBB تقریباً نزدیک‌تر به مقدار واقعی توزیع نرمال آماره T_n است.



شکل ۲- توزیع واقعی آماره و برآوردهای بوت‌استرپ آن

بحث و نتیجه‌گیری

یکی از روش‌های بوت‌استرپ مورد استفاده در داده‌های فضایی بلوک متحرک است، که در آن مشاهدات را در بلوک‌هایی متحرک در نظر گرفته و از آن‌ها باز نمونه‌گیری می‌شود. در این روش مشاهدات مرزی ناحیه نمونه‌گیری نسبت به مشاهدات مرکزی امکان کمتری را برای حضور در بلوک‌ها دارند و این مسئله باعث ایجاد حالت اریبی در برآوردها می‌شود.

در این مقاله برای رفع این مشکل، روش بوت‌استرپ بلوک مجزا ارائه شد، که در آن مشاهدات به بلوک‌های مجزا افراز و باز نمونه‌گیری از این بلوک‌ها انجام می‌شود. نداشتن خطا در برآورد اریبی برآوردگر میانگین نمونه‌ای به روش بوت‌استرپ بلوک مجزا و سازگاری برآوردگر واریانس نشان داده شد. همچنین در یک بررسی شبیه‌سازی نشان داده شد که معیار MSE برای برآورد اریبی به روش بوت‌استرپ بلوک مجزا نسبت به روش بلوک متحرک کاهش چشم‌گیری دارد، در حالی‌که برآورد واریانس به هر دو روش بلوک مجزا و متحرک دقت تقریباً یکسانی دارند. همچنین تأثیر اندازه نمونه و خاصیت سازگاری برآوردگرهای اریبی و واریانس به هر دو روش مورد بررسی شبیه‌سازی قرار گرفت. علت اصلی افزایش دقت برآورد اریبی شانس یک‌سان ظاهر شدن مشاهدات در بلوک‌های مختلف و رفع مشکل نقاط مرزی است. نظر به این‌که فرض محدود کننده مانایی میدان تصادفی ممکن است در مسایل کاربردی محقق نباشد، یا برآوردگرهایی غیر از میانگین مورد

نیاز باشند، می‌توان نحوه اجرا و میزان دقت ارائه شده را برای میدان‌های تصادفی نامانا مورد بررسی بیشتر قرار داد.

منابع

1. N. Cressie, *Statistics for Spatial Data*, John Wiley, New York (1993).
2. B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, *Annals of Statistics*, 7 (1979) 1-26.
3. P. Hall, *Resampling a Coverage Pattern*, *Stochastic Processes and their Application*, 20 (1985) 231-246.
4. P. Hall, *On Confidence Intervals for Spatial Parameters Estimated from Nonreplicated Data*, *Biometrika*, 44 (1988) 271-277.
5. P. Buhlmann and H. R. Kunsch, *Comments on "Prediction of Spatial Cumulative Distribution Functions Using Subsampling"*, *Journal of the American Statistical Association*, 94 (1999) 97-99.
6. S.N. Lahiri, *Resampling Methods for Dependent Data*, Springer-Verlag, New York (2003).
7. P. Doukhan, *Mixing: Properties and Examples*, vol. 85 of *Lecture Notes in Statistics*, Springer-Verlag, New York (1994).