روش بوت‌استرپ بلوک مجزا در آمار فضایی

نصراله ایران‌پناه و محسن محمدزاده: دانشگاه تربیت مدرس

چکیده

روش بوت‌استرپ افرون برای برآورد میزان دقت برآوردگرها، هنگام مشاهدات مستقل کاربرد دارد. برای داده های فضایِی که بر حسب موقعیت قرار گرفتن آن‌ها در فضای مورد بررسی به یک‌دیگر وابسته هستند، معمولاً روش بوت‌استرپ بلوک متحرک مورد استفاده قرار می‌گیرد. از آنجا که در این روش مشاهدات مرزی نسبت به سایر مشاهدات امکان کمتری برای حضور در بلوک‌ها دارند، در این مقاله روش بوت‌استرپ بلوك مجزا معرفی و الگوریتمی برای برآورد میزان دقت برآوردگرها ارائه می‌شود. همچنین کارایی روش بوت‌استرپ بلوك مجزا با بلوك متحرك در يك بررسی شبیه‌سازی مورد مقایسه عددی قرار گرفته، نشان داده می‌شود برآوردگر اریبی میانگین نمونه‌ای به روش بوت‌استرپ بلوك مجزا بدون خطا و برآوردگر واریانس آن سازگار است.

مقدمه

در آمار کلاسیک عموماً فرض می‌شود داده‌های به دست آمده از جامعه، مستقل‌اند. در عمل با موارد زیادی مواجه می‌شویم که داده‌ها وابسته‌اند. داده‌های فضایی[١] مشاهداتی هستند که وابستگی آن‌ها ناشی از موقعیت‌شان در فضای بررسی شده است و این وابستگی تابعی از فاصله مشاهدات از یک‌دیگر است. اغلب استنباط‌های آمار فضایِی مبتنِی بر گاوسِی بودن میدان تصادفِی است، که در عمل ممکن است این شرط برقرار نباشد. در این‌گونه موارد می‌توان از الگوریتم بوت‌استرپ و بازنمونه‌گیری[١] از داده‌های فضایِی در استنباط استفاده کرد.

افرون[٢] روش بوت‌استرپ را برای داده‌های مستقل ارائه کرد؛ در این روش می‌توان با استفاده از بازنمونه‌گیری داده‌ها، اریبی، واریانس و توزیع برآوردگرها را برآورد کرد. این روش برای داده‌های فضایِی به علت وابستگی مشاهدات کاربرد ندارد. هال[٣] دو روش براساس بلوك‌ی کردن مشاهدات و موقعیت‌ها برای حالت خاص داده‌های موزاییك ارائه کرد، که در آن‌ها همبستگی مشاهدات در موقعیت مجاور مرز بلوک‌ها ساختار واقعی ندارد، به این علت هال[٤] کوشیده است این مشکل را با استفاده از نامساوی بن فرونی به نوعی حل کند. بولمان و کونش[٥] و لاهیری[٦] نیز روش بوت‌استرپ بلوك متحرك[٢](MBB) را برای داده‌های فضایِی ارائه کردند، که در آن ساختار داده‌ها در فضای $\mathbb{Z}^d$ مورد توجه قرار گرفته است. در این روش نمونه‌ای از مشاهدات

واژه‌های کلیدی: بوت‌استرپ بلوك مجزا، بوت استرپ بلوك متحرك، α - اختلاط، شبیه‌سازی مونت کارلو

۱-Resampling ۲-Moving Block Bootstrap

در بلوك‌های متحرك بازنمونه‌گیری می‌شود، به گونه‌ای که هر مشاهده حداقل در یكی از بلوك‌ها قرار گیرد، اما امکان کم‌تر مشاهدات مرزی در مقایسه با مشاهدات مرکزی برای حضور در بلوك‌ها موجب اریبی برآوردگرها می‌شود.

در این مقاله برای رفع این مشکل روش بوت‌استرپ بلوك 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 بلوك‌های مجزا اجرا می‌شود. برای این منظور، الگوریتم بوت‌استرپ IID در بخش ۲ ارائه می‌شود. در بخش ۳ روش بوت‌استرپ بلوك مجزا ارائه و روش بوت‌استرپ بلوك متحرك معرفی می‌شود. در بخش ۴ خواص برآوردگرهای اریبی و واریانس میانگین نمونه‌ای به روش بوت‌استرپ بلوك مجزا مورد بررسی قرار می‌گیرد. نتایج حاصل از دو روش بلوك مجزا و متحرك با استفاده از تکنیك 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 $\mathcal{Z} \equiv \{Z_1,\ldots,Z_n\}$ مجموعه‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع تجمعی F_θ و کمیت تصادفی مورد نظر $T = t(\mathcal{Z};F_\theta)$ 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 برآوردگر پارامتر مورد نظر θ و بر اساس مشاهدات مستقل به صورت زیر پیشنهاد کرد:

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

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

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

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

$$Bias_*(T^*) = E_*(T^*) - T \text{ ، } Var_*(T^*) = E_*\left[T^* - E_*(T^*)\right]^2 \text{ و } G_*(t) = P(T^* \le t)$$

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

استفاده از

۱-Separate Block Bootstrap

شبیه‌سازی مونت کارلو و تکرار B بار مراحل ب و ج ومحاسبه $T_1^*,\ldots,T_B^*$ به ترتیب به صورت:

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

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

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

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

در اغلب تحلیل‌های نظری بوت‌استرپ برای F_θ و T شرایطی منظور می‌شود، به طوری که با افزایش n توزیع و گشتاورهای T^* به توزیع و گشتاورهای مجانبی T میل کند.

بوت‌استرپ فضایی

معمولاً میدان تصادفی $\{Z(s):s\in D\}$ به عنوان مدل آماری برای تجزیه و تحلیل داده‌های فضايی در نظر گرفته می‌شود، که در آن D یك مجموعه اندیس‌گذار در فضای اقلیدسی $d\geq 1$ بعدی $\mathbb{R}^d$ است. فرض کنید ناحیه نمونه‌گیری D_n وقتی که $n\to\infty$ غیرکران‌دار باشد. گیریم $\tilde{D}\subset\left[-\frac{1}{2},\frac{1}{2}\right)^d$ یك مجموعه باز هم‌بند شامل مبدأ و D_o یك مجموعه نمونه اولیه برای نواحی نمونه‌گیری با شرط $\tilde{D}\subset D_o\subset cl.\left(\tilde{D}\right)$ باشد، که در آن $cl.\left(\tilde{D}\right)$ بستار مجموعه $\tilde{D}$ است. همچنین با فرض آن‌که $\{\lambda_n\}_{n\geq 1}$ یك دنباله صعودی از اعداد حقیقی بزرگ‌تر از یك باشد، ناحیه نمونه‌گیری D_n با تکثیر مجموعه نمونه اولیه D_o و ثابت مقیاس‌بندی λ_n به صورت $D_n=\lambda_n D_o$ در نظر گرفته می‌شود. فرض کنید در ناحیه نمونه‌گیری D_n، تعداد N_n موقعیت محدود $S_n\equiv\{s_1,\ldots,s_{N_n}\}$ از یك میدان تصادفی مانای $\{Z(s):s\in\mathbb{Z}^d\}$ به صورت یك شبکه منظم مشاهده شده است. اندازه نمونه N_n و حجم ناحیه نمونه‌گیری D_n به صورت:

$$N_n=Vol.\left(D_o\right)\cdot\lambda_n^d \qquad (۱)$$

با هم در ارتباط هستند. با قرار دادن $N=N_n$، فرض کنید $T_n=t_n\left(\mathcal{Z}_n\right)$ برآوردگر پارامتر θ باشد، که در آن $\mathcal{Z}_n=\left(Z\left(s_1\right),\ldots,Z\left(s_N\right)\right)$ یك نمونه از مشاهدات است. هدف تعیین مشخصات توزیع نمونه‌ای T_n به

روش بوت‌استرپ بلوك فضايی است. در اين بخش روش جديد بوت‌استرپ بلوك مجزا معرفی شده و با استفاده از نمادهای آن روش بلوك متحرك نيز بيان می‌شود.

۱ـ بوت‌استرپ بلوك مجزا

فرض كنيد $\{\beta_n\}_{n\geq 1}$ يك دنباله از اعداد صحيح مثبت باشد، به طوری كه:

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

يعنی β_n با نرخ آهسته‌تری از عامل مقياس‌بندی λ_n برای ناحيه نمونه‌گيری D_n به بی‌نهايت ميل كند. β_n عامل مقياس‌بندی برای بلوك‌ها يا زير نواحی در روش بوت‌استرپ بلوك فضايی است. ابتدا ناحيه نمونه‌گيری D_n را با استفاده از مكعب‌هايی به حجم β_n^d افراز می‌كنيم. گيريم $K_n=\{k\in\mathbb{Z}^d:\beta_n(k+U)\cap D_n\neq\phi\}$ مجموعه انديس مكعب‌هايی به شكل $\beta_n(k+U)$ باشد كه اشتراك غير تهی با ناحيه نمونه‌گيری D_n دارد و در آن $U=[0,1)^d$ مكعب واحد در $\mathbb{R}^d$ است. با فرض $N=\beta_n^d|K_n|$، ناحيه نمونه‌گيری D_n با مكعب‌هايی به حجم β_n^d و تعداد $|K_n|$ كامل می‌شود. سپس برای هر $k\in K_n$، نمونه بوت‌استرپ را در زير ناحيه kام به صورت:

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

در نظر می‌گيريم $I_n=\{i\in\mathbb{Z}^d:\beta_n(i+U)\subset D_n\}$ مجموعه انديس مكعب‌هايی مجزا به حجم β_n^d در D_n با نقاط شروع $i\in\mathbb{Z}^d$ باشد، در نتيجه $\{\beta_n(i+U):i\in I_n\}$ يك مجموعه از زير نواحی يا بلوك‌های مكعبی مجزا واقع در D_n هستند. فرض كنيد $\mathcal{Z}_n(D_n)=\{Z(s_1),\ldots,Z(s_N)\}$ نمونه كامل و $\mathcal{Z}_n(D_n(k))$ زير نمونه قرار گرفته در زير ناحيه $D_n(k)$ باشد. برای به دست آوردن يك نمونه بوت‌استرپ بلوك مجزای فضايی، ابتدا برای هر $k\in K_n$ يك بلوك به صورت تصادفی از مجموعه $\{\beta_n(i+U):i\in I_n\}$ و مستقل از بلوك‌های ديگر انتخاب می‌كنيم. سپس با استفاده از مشاهدات در زير ناحيه بازنمونه‌گيری شده، نمونه بوت‌استرپ را روی زير ناحيه $D_n(k)$ تعيين می‌كنيم. سرانجام، فرض كنيد $K=K_n$ اندازه مجموعه K_n و همچنين $\{I_k:k\in K_n\}$ يك مجموعه از K متغير iid با توزيع مشترك

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

باشد. برای هر $k\in K_n$ زير نمونه بوت‌استرپ بلوك مجزای $\mathcal{Z}_n^*(D_n(k))$ را با استفاده از بلوك بازنمونه‌گيری شده $\mathcal{Z}_n(\beta_n[I_k+U])$ به صورت:

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

به دست می‌آوریم. حال نمونه بوت‌استرپ بلوك مجزا $Z_n^*\left(D_n\right)$ را از به هم پیوستن بلوك‌های بازنمونه‌گیری شده $\left\{Z_n^*\left(D_n\left(k\right)\right):k\in K_n\right\}$ تعیین و نسخه بوت‌استرپ بلوك مجزای آماره $T_n=t_n\left(Z_n\right)$ را به صورت $T_n^*=t_n\left(Z_n^*\left(D_n\right)\right)$ تعریف می‌کنیم. در واقع نمونه بوت‌استرپ بلوك مجزای $Z_n^*\left(D_n\right)$ يك نمونه تصادفی ساده باجای‌گذاری به حجم K از بلوك‌های مکعبی مجزای $Z_n^*\left(D_n\left(k\right)\right)$ به حجم β_n^d است. ادامه روش مانند مرحله (د) بوت‌استرپ IID است.

۲- بوت‌استرپ بلوك متحرك

بولمان و کونش[۵] و لاهیری[٦] روش بوت‌استرپ بلوك متحرك یا متداخل را برای برآورد مشخصات توزیع T_n ارائه کردند. فرض کنید زیر نواحی $D_n\left(k\right)$ افراز ناحیه نمونه‌گیری D_n داده شده در(۳) به صورت مکعب‌هايی به حجم β_n^d باشد. در این روش ابتدا با استفاده از مجموعه مکعب‌های $J_n=\left\{j\in\mathbb{Z}^d:\left(j+\beta_nU\right)\subset D_n\right\}$ كه متداخل هستند، K متغیر تصادفی iid $\left\{J_k:k\in K_n\right\}$ با توزیع مشترك $P\left(J_1=j\right)=\frac{1}{|J_n|},\quad j\in J_n$ تولید می‌شود. سپس زیر نمونه بوت‌استرپ بلوك متحرك به صورت:

$$Z_n^{**}\left(D_n\left(k\right)\right)=Z_n\left(\left[J_k+\beta_nU\right]\cap\left[D_n\left(k\right)-k\,\beta_n+J_k\right]\right)$$

به دست می‌آید. حال نسخه بوت‌استرپ بلوك متحرك آماره T_n به صورت $T_n^{**}=t_n\left(Z_n^{**}\left(D_n\right)\right)$ تعریف می‌شود، كه در آن $Z_n^{**}\left(D_n\right)$ از به‌هم پیوستن بلوك‌های بازنمونه‌گیری شده $Z_n^{**}\left(D_n\left(k\right)\right)$ تعیین می‌شوند. در واقع نمونه بوت‌استرپ بلوك متحرك $Z_n^{**}\left(D_n\right)$ يك نمونه تصادفی ساده باجای‌گذاری به حجم K از بلوك‌های مکعبی متداخل $Z_n^{**}\left(D_n\left(k\right)\right)$ به حجم β_n^d است.

خواص برآوردگرهای بوت‌استرپ اریبی و واریانس میانگین نمونه‌ای

فرض کنید برای میدان تصادفی مانای $\left\{Z\left(i\right):i\in\mathbb{Z}^d\right\}$، میانگین نمونه‌ای $\bar{Z}_n=N^{-1}\sum_{i=1}^{N}Z\left(s_i\right)$ به عنوان برآوردگر میانگین میدان تصادفی $\mu=E\left[Z\left(\cdot\right)\right]$، براساس مشاهدات Z_n باشد. اگر $\bar{Z}_n^*$ میانگین نمونه‌ای بوت‌استرپ بلوك مجزا باشد، در این صورت برآورد بوت‌استرپ پارامتر $\mathrm{Bias}\left(\bar{Z}_n\right)=E\left(\bar{Z}_n\right)-\mu=0$ به صورت $\widehat{\mathrm{Bias}}\left(\bar{Z}_n\right)=\mathrm{Bias}_*\left(\bar{Z}_n^*\right)=E_*\left(\bar{Z}_n^*\right)-\bar{Z}_n$ تعیین می‌شود.

لم۱: روش بوت‌استرپ بلوك مجزا، اریبی میانگین نمونه‌ای از يك میدان تصادفی مانای $\left\{Z\left(i\right):i\in\mathbb{Z}^d\right\}$ را بدون خطا برآورد می‌کند.

اثبات: مجموع‌های $S_n\left(i;k\right)=\sum_{j\in B_n\left(i;k\right)}Z\left(j\right)$ و $S_n^*\left(k\right)=\sum_{j\in B_n\left(I_k;k\right)}Z^*\left(j\right)$، که در آن‌ها $B_n\left(i;k\right)\equiv D_n\left(k\right)-k\,\beta_n+i\,\beta_n$، iامین بلوك نوع k است $\left(k\in K_n,i\in I_n\right)$ را در نظر بگیرید. به‌دلیل شانس حضور یك‌سان مشاهدات در بلوك‌های مجزای $\left\{\beta_n\left(i+U\right):i\in I_n\right\}$ و با توجه به $\left|K_n\right|=\left|I_n\right|$ داریم:

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

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

در روش MBB برخلاف روش SBB به دلیل امکان نابرابر مشاهدات برای حضور در بلوك‌های متحرك $\left\{\left(j+\beta_nU\right):j\in J_n\right\}$ برآورد اریبی میانگین نمونه‌ای همراه با خطاست. برای رفع این مشکل معمولاً با مرکزی کردن آماره به شکل $T_n=\bar{Z}_n-\mu$ برآورد بوت‌استرپ آن به صورت $T_n^*=\bar{Z}_n^*-E_*\left(\bar{Z}_n^*\right)$ در نظر گرفته می‌شود.

برآورد بوت‌استرپ پارامتر $\sigma_n^2=N\mathrm{Var}\left(\bar{Z}_n\right)$ را می‌توان بر اساس پارامتر اندازه- بلوك β_n به صورت

$$\hat{\sigma}_n^2=\hat{\sigma}_n^2\left(\beta_n\right)=N\mathrm{Var}_*\left(\bar{Z}_n^*\right)\qquad(۴)$$

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 $\sigma-$جبر مستقل $\boldsymbol{A}$ و $\boldsymbol{B}$ از $\boldsymbol{F}$ در یك فضای احتمال $\left(\Omega,\boldsymbol{F},P\right)$ برای هر $A\in\boldsymbol{A}$ و $B\in\boldsymbol{B}$، $\Delta\equiv P\left(A\cap B\right)-P\left(A\right)P\left(B\right)=0$ است. اما وقتی $\boldsymbol{A}$ و $\boldsymbol{B}$ مستقل نباشند، معمولاً اندازه وابستگی $\boldsymbol{A}$ و $\boldsymbol{B}$ با بزرگترین مقدار وابستگی Δ به صورت:

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

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

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

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

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

فاصله مجموعه‌های S_1 و S_2 است و

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

است، به طوری که $\boldsymbol{F}_Z(S)$، $\sigma-$جبر تولید شده توسط متغیر تصادفی $\left\{Z(s):s\in S\cap \mathbb{Z}^d\right\}$ است.

لم ۲: میدان تصادفی $\left\{Y(i):i\in \mathbb{Z}^d\right\}$ را با فرض $E\left[Y(i)\right]=0$ و $E\left|Y(i)\right|^{2q+\delta}<\infty$ برای هر $i\in \mathbb{Z}^d$ و عددی طبیعی مانند q و $\delta\in(0,\infty)$ در نظر بگیرید. فرض کنید

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

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

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

که در آن ثابت $C(q,\delta)$ تنها به q، δ و d و ضریب $\alpha-$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.

همان‌طور که لاهیری[٦] سازگاری برآوردگر واریانس میانگین نمونه‌ای به روش بوت‌استرپ بلوک متحرك را نشان داد، این خاصیت برای روش بوت‌استرپ بلوك مجزا نیز در قضیه زیر اثبات می‌شود.

قضیه ۱: فرض کنید میدان تصادفی $\left\{Z(i):i\in \mathbb{Z}^d\right\}$ مانا با $E\left|Z(\cdot)\right|^{6+\delta}<\infty$ و ضریب $\alpha-$اختلاط $\alpha(a;b)\le ca^{-\tau_1}b^{\tau_2}$ برای $a\ge1$ و$b\ge1$ و به ازای عددی نامنفی مانندδ، $\tau_1\ge\dfrac{5d(6+\delta)}{\delta}$ و $0\le\tau_2\le\dfrac{\tau_1}{d}$ باشد. اگر $\beta_n^{-1}+\lambda_n^{-1}\beta_n=o(1)$ آنگاه

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

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

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

اثبات : بدون از دست دادن کلیت مسئله فرض می‌شود $\mu=0$. با توجه به استقلال بلوك‌های بازنمونه‌گیری شده و $N=\beta_n^d\left|K_n\right|$ داریم

$$\begin{aligned} N\mathrm{Var}_*\left(\bar{Z}_n^*\right) &= N\mathrm{Var}_*\left(N^{-1}\sum_{k\in K_n} S_n^*(k)\right) \\ &= N^{-1}\sum_{k\in K_n}\mathrm{Var}_*\left[S_n^*(k)\right] \\ &= \beta_n^{-d}\left[E_*\left(S_n^*(\cdot)\right)^2-\left(E_*\left(S_n^*(\cdot)\right)\right)^2\right] \\ &\equiv \hat{Q}_{1n}-\hat{Q}_{2n} \end{aligned} \tag{۵}$$

همچنین داریم

$$\begin{aligned} E\left(\hat{Q}_{1n}\right) &= E\left[\beta_n^{-d}E_*\left(S_n^*(\cdot)\right)^2\right] \\ &= \beta_n^{-d}E\left[S_n(i;\cdot)\right]^2 \\ &= \beta_n^{-d}E\left[\sum_{j\in\beta_n U\cap\mathbb{Z}^d} Z(j)\right]^2 \\ &= \sum_{j\in\beta_n U\cap\mathbb{Z}^d} E\left[Z(0)Z(j)\right]\to\sigma_\infty^2 \quad as \quad n\to\infty \end{aligned} \tag{۶}$$

با توجه به لم ۱ داریم

$$\begin{aligned} E\left(\hat{Q}_{2n}\right) &= E\left[\beta_n^{-d}\left(E_*\left(S_n^*(\cdot)\right)\right)^2\right] \\ &= \beta_n^{-d}E\left[N\left|K_n\right|^{-1}\bar{Z}_n\right]^2 \\ &= \beta_n^{d}\mathrm{Var}\left[\hat{Z}_n(s_0)\right] \\ &= O\left(\lambda_n^{-d}\beta_n^{d}\right) \end{aligned} \tag{۷}$$

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

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

طور مشابه داریم ه ب

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

بنا بر روابط (۸) و(۹) وقت‌ی $n \to \infty$، آنگاه $\operatorname{Var}\left(\hat{Q}_{1n}\right) \to 0$ و $\operatorname{Var}\left(\hat{Q}_{2n}\right) \to 0$. بنا بر این با توجه به نااریب‌ی 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 $\alpha-$ اختلاط در دخان[۷] مورد بررس‌ی قرار گرفته است.

مقایسه بوت‌استرپ بلوك مجزا و متحرك

در این بخش، دو روش SBB و MBB با استفاده از تکنیك شبیه‌سازی مونت کارلو برای داده‌های فضاي‌ی (SMS) مورد مقایسه عددی قرار می‌گیرند. فرض کنید $\left\{Z(s): s \in \mathbb{Z}^2\right\}$ يك میدان تصادف‌ی گاوس‌ی مانا‌ی مرتبه دوم با میانگین صفر و هم تغییرنگار نماي‌ی

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

باشد. با فرض $\theta=(c_0,c,a)=(1,1,1)$ و داده‌های فضايی در يك شبكه منظم مستطيلی 20×30 $(N=600)$ قرار داشته باشند، با در نظر گرفتن فاصله اقليدسی بين دو موقعيت، ماتريس كواريانس ميدان تصادفی به صورت $\Sigma=\left(\sigma\left(s_i-s_j\right)\right)$ به دست می‌آيد. اگر $\Sigma^{1/2}$ تجزيه چولسكی[1] ماتريس Σ به صورت $\Sigma=\Sigma^{1/2}\times\Sigma^{1/2}$ باشد، يك نمونه تصادفی $Z_n=\left(Z\left(s_1\right),\ldots,Z\left(s_N\right)\right)$ از ميدان تصادفی را می‌توان به صورت $Z_n=\Sigma^{1/2}X_n$ به دست آورد. كه درآن $X_n=\left(X_1,\ldots,X_N\right)$ يك نمونه تصادفی(iid) از توزيع $N\left(0,1\right)$ است. اگر $T_n=t_n\left(Z_n\right)$ برآوردگر پارامتر θ باشد، با تكرار B بار (به قدر كافی بزرگ) نمونه‌گيری به حجم N و محاسبه مقادير $T_1,\ldots,T_B$ می‌توان اريبی، واريانس و توزيع $T_n=\sqrt{N}\,\bar{Z}_n$ را با هر اندازه دقت‌ی به ترتيب به صورت

$$Bias\left(T_n\right)\simeq\bar{T}-\theta$$

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

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

تقريب كرد، كه در آن‌ها $\bar{T}=\frac{1}{B}\sum_{i=1}^{B}T_i$ است. مقادير واقعی اريبی و واريانس آماره $T_n=\sqrt{N}\,\bar{Z}_n$ كه در آن $\bar{Z}_n$ ميانگين نمونه‌ای به عنوان برآوردگر پارامتر $\theta=\sqrt{N}\,\mu$ است، برای 10000 بار تكرار نمونه‌گيری به حجم $N=600$ محاسبه و در ستون SMS جدول ۱ نشان داده شده است. مقادير برآورد اريبی و

[1]-Choleski Decomposition

واريانس آماره T_n به دو روش بوت‌استرپ بلوك مجزا و متحرك به ترتيب در دو ستون SBB و MBB جدول ۱ نشان داده شده است. برای انجام روش بوت‌استرپ بلوك‌ی فضايی، ناحيه نمونه‌گيری را به صورت $D_n=[-10,10)\times[-15,15)$ با ثابت مقياس‌بندی $\lambda_n=30$، اندازه بلوك $\beta_n=5$ و مجموعه نمونه اوليه $D_0=\left[-\frac{1}{3},\frac{1}{3}\right)\times\left[-\frac{1}{2},\frac{1}{2}\right)\subset\left[-\frac{1}{2},\frac{1}{2}\right)^2$ درنظر می‌گيريم. در اين دو روش كه با استفاده از مجموعه مشاهده

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

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

افراز می‌کنیم. در روش SBB، بازنمونه‌گیری از بلوک‌های مجزای

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

و در روش MBB، بازنمونه‌گیری از بلوک‌های متحرک

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

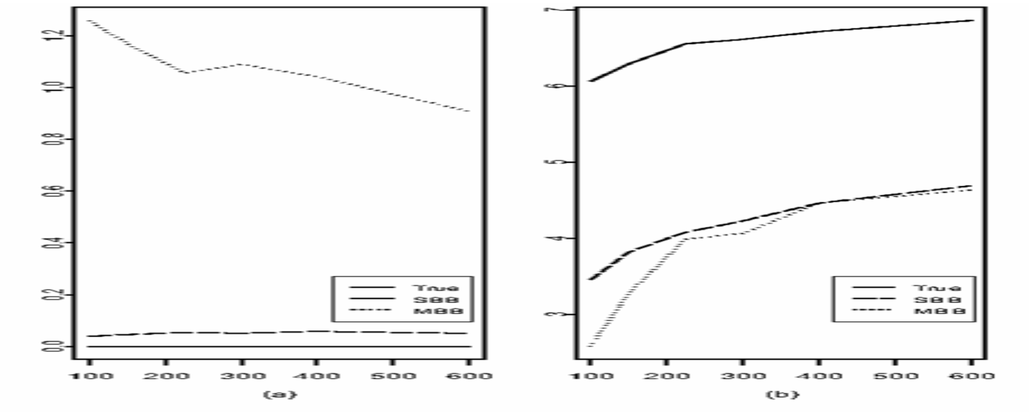
انجام می‌شود. تعداد کل بلوک‌های مجزا و متحرک در این شبکه به ترتیب 24 و 416 است که از بین آن‌ها به روش نمونه‌گیری تصادفی ساده با جایگذاری، 24 بلوک به حجم 25 را انتخاب و از به هم پیوستن آن‌ها یک نمونه بوت‌استرپ $Z_n^*=\{Z^*(s_1),...,Z^*(s_N)\}$ به حجم $N=600$ تولید می‌شود. با محاسبه آماره بوت‌استرپ $T_n^*=\sqrt{N}\,\bar{Z}_n^*$ و تکرار 2000 بار این الگوریتم، $T_1^*,...,T_{2000}^*$ به دست می آید. 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	٦/٨٦	٤/٧٨	٤/٦٧

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 برآوردگرهای 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 ۱(a) ملاحظه می شود، برآورد اریبی به روش SBB تقریباً

بدون خطا، نزدیک مقدار واقعی اریبی است. در حالی‌که برآورد اریبی به روش MBB همراه با خطاست، ولی با افزایش اندازه نمونه، به مقدار واقعی صفر نزدیك 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 نزدیك 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 ملاك میانگین مربعات خطا (MSE) حاصل از 1000 بار تکرار برآورد بوت‌استرپ اریبی و واریانس آماره T_n به صورت

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

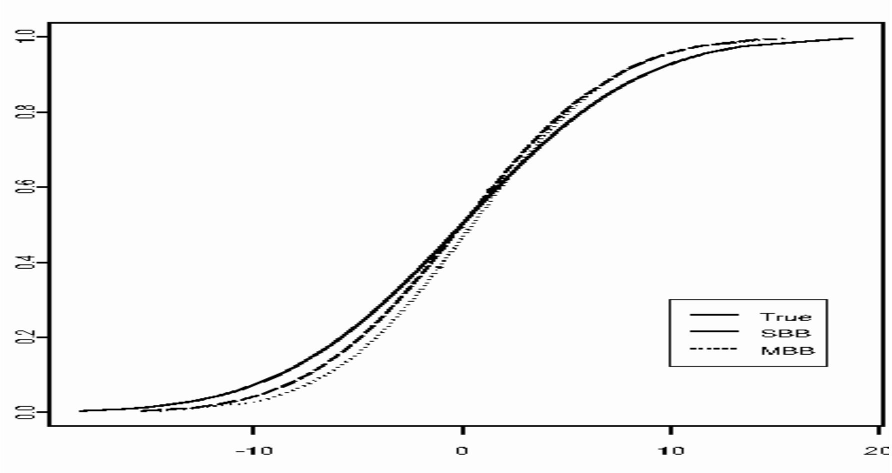
استفاده شده است، که در آن R مقدار واقعی اندازه دقت آماره T_n است و $\hat{R}_*(i)$ نشان دهنده برآورد بوت‌استرپ این اندازه‌ها در تکرار iام است. با مقایسه مقادیر MSE برای دو روش SBB و MBB که در جدول ۲ نشان داده شده‌اند، ملاحظه می‌شود که برآورد اریبی به روش SBB نسبت به MBB از دقت بسیار بیش‌تری برخوردار است،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

جدول ۲- مقادیر MSE برآوردهای اریبی و واریانس

اندازه دقت	SBB	MBB
اریبی	۰/۰۰٤	۱/۲۷۳

واریانس	٦/٥١	٦/٩٥

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 بازنمونه‌گیری می شود. 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 برآوردگرها می‌شود.

در این مقاله برای رفع این مشکل، روش بوت‌استرپ بلوک مجزا ارائه شد، که در آن مشاهدات به بلوک‌های مجزا افراز و بازنمونه‌گیری 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 Caverage 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).