برآورد اندازه‌های دقت کریگیدن به روش خودگردانی بلوکی فضایی

نصراله ایران‌پناه، محسن محمدزاده*

گروه آمار، دانشکده علوم پایه، دانشگاه تربیت مدرس، تهران، ایران

* مسئول مکاتبات- آدرس الکترونیکی: *mohsen_m@modares.ac.ir*

(دریافت: ۸۴/۱۰/۱۹ ؛ پذیرش:۸۵/۸/۲۴)

چکیده

برای داده‌های فضایی که بر حسب موقعیت قرار گرفتن آنها در فضای مورد مطالعه به یکدیگر وابسته‌اند، معمولاً روش خودگردانی «بلوک متحرک» به منظور برآورد اندازه‌های دقت برآوردگرها استفاده می‌شود. چون در این روش حضور مشاهدات مرزی در بلوکهای بازنمونه‌گیری شده نسبت به سایر مشاهدات شانس کمتری دارند، برآوردگرهای اندازه‌های دقت اریب خواهند بود. در این مقاله الگوریتم خودگردانی «بلوک مجزا» برای برآورد اندازه‌های دقت پیشگوی فضایی کریگیدن ارائه می‌شود. سپس نشان داده می‌شود برآورد اریبی کریگیدن به روش خودگردانی بلوک مجزا نااریب و برآوردگر واریانس کریگیدن سازگار است. نهایتاً در یک مطالعه شبیه‌سازی کارایی روش خودگردانی بلوک مجزا در برآورد اندازه‌های دقت با روش خودگردانی بلوک متحرک مورد مقایسه قرار می‌گیرد.

واژه‌های کلیدی: کریگیدن، خودگردانی بلوک مجزا، خودگردانی بلوک متحرک

مقدمه

اغلب، استنباط آمار فضایی مبتنی بر گاوسی بودن میدان تصادفی است، که در عمل ممکن است محقق نباشد. در اینگونه موارد الگوریتم خودگردان و بازنمونه‌گیری (Resampling) از داده‌های فضایی را می‌توان برای استنباط داده‌های فضایی بکار گرفت. اما روش خودگردان توسط افرون (Efron 1979) برای داده‌های مستقل ارائه گردیده است، که با استفاده از بازنمونه‌گیری داده‌ها، اندازه‌های دقت اریبی و واریانس و توزیع برآوردگرها برآورد می‌شوند. این روش برای داده‌های فضایی به علت وابستگی مشاهدات کاربرد ندارد. هال (Hall, 1985, 1988) دو روش براساس بلوکی کردن مشاهدات و موقعیتها برای حالت خاص داده‌های موزاییک ارائه کرد. سوستت‌دلونا و یانگ (Sjostedt-DeLuna & Young, 2003) با فرض گاوسی بودن میدان تصادفی یک فاصله پیشگوی کالبیده (Calibrated) براساس خودگردان ارائه کردند. بولمان و کونش (Buhlmann & Kunsch, 1999) و ژو و لاهیری (Zhu & Lahiri, 2001) نیز روش خودگردانی بلوک متحرک (MBB: Moving Block Bootstrap) را برای داده‌های فضایی ارائه کردند، که در آن ساختار داده‌ها در زیر فضای مشبکه‌ای منظم از $\mathbb{Z}^d$ مورد توجه قرار گرفته است. در این روش نمونه‌ای از مشاهدات در بلوکهای متحرک بازنمونه‌گیری می‌شود، بگونه‌ای که هر مشاهده حداقل در یکی از بلوکها قرار گیرد. شانس کمتر مشاهدات مرزی در مقایسه با مشاهدات مرکزی برای حضور در بلوکها موجب اریبی برآوردگرهای این روش می‌شود. برای رفع این مشکل ایران‌پناه و محمدزاده (۱۳۸۴) روش خودگردانی بلوک مجزا (SBB: Separate Block Bootstrap) را برای برآورد اندازه‌های دقت میانگین داده‌های فضایی معرفی نمودند، که در آن الگوریتم خودگردان با استفاده از بازنمونه‌گیری بلوکهای افراز شده، اجرا می‌شود.

در این مقاله برآورد اندازه‌های دقت پیشگوی فضایی کریگیدن (Kriging) به روش SBB معرفی شده است. برای این منظور روشهای SBB و MBB در بخش ۲ معرفی می‌شوند. در بخش ۳ خواص برآوردگرهای اریبی و واریانس کریگیدن به روش SBB مورد بررسی قرار می‌گیرد. نتایج حاصل از دو روش SBB و MBB با استفاده از تکنیک شبیه‌سازی در بخش ۴ مورد مقایسه عددی قرار گرفته و در بخش آخر به بحث و نتیجه‌گیری پرداخته شده است.

خودگردانی بلوکی فضایی

داده‌های فضایی مشاهداتی هستند که وابستگی آنها ناشی از موقعیتشان در فضای مورد مطالعه است و این وابستگی تابعی از فاصله مشاهدات از یکدیگر است. معمولاً برای مدل‌بندی داده‌های فضایی از میدان تصادفی $\{Z(s): s\in D\}$ استفاده می‌شود، که در آن D یک مجموعه اندیس گذار درفضای اقلیدسی $d\geq 1$ بعدی $\mathbb{R}^d$ است. فرض کنید $\boldsymbol{Z}=\left(Z(s_1),\ldots,Z(s_N)\right)$ بردار مشاهدات از میدان تصادفی مانای $\{Z(s): s\in D\}$ با میانگین $\mu=E\left[Z(\cdot)\right]$ و تابع

همتغییرنگار $\sigma(h)=\mathrm{Cov}\left[Z(\cdot),Z(h)\right]$ باشد. هدف پیشگویی مقدار میدان تصادفی در موقعیت جدید $s_.$ است. ماترون (Matheron, 1963) بهترین پیشگوی خطی ناریب برای $Z(s_.)$ را با کمینه کردن میانگین توان دوم خطای پیشگویی $E\left[Z(s_.)-\sum_{i=1}^{N}\lambda_i Z(s_i)\right]^2$ بصورت $\hat{Z}(s_.)=\lambda'\boldsymbol{Z}$ معرفی نمود، که بردار ضرایب λ بصورت

$$\lambda'=(c+\mathbf{1}m)'\Sigma^{-1},\qquad m=\frac{1-\mathbf{1}'\Sigma^{-1}c}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \qquad (1)$$

محاسبه می‌شود و در آن $\mathbf{1}$ یک بردار $N\times 1$ از اعداد ۱، c یک بردار $N\times 1$ با iامین عنصر $\sigma(s_.-s_i)$ و Σ یک ماتریس $N\times N$ با (i,j)امین عنصر $\sigma(s_i-s_j)$ است. ماترون (Matheron 1963) این پیشگوی بهینه را کریگیدن نامید که واریانس آن بصورت $\sigma_k^2(s_.)=\sigma(\cdot)-\lambda'c+m$ بدست می‌آید. در این مقاله، هدف تعیین اندازه‌های دقت کریگیدن $\hat{Z}(s_.)=\sum_{i=1}^{N}\lambda_i Z(s_i)$ یا عبارت $T=\hat{Z}(s_.)-Z(s_.)$ به روش خودگردانی بلوکی فضایی است.

فرض کنید $\tilde{D}\subset\left[-\frac{1}{2},\frac{1}{2}\right]^d$ یک مجموعه باز همبند شامل مبدأ و $D_.$ یک مجموعه نمونه اولیه با شرط $\tilde{D}\subset D_.\subset cl.(\tilde{D})$ باشد، که در آن $cl.(\tilde{D})$ بستار مجموعه $\tilde{D}$ است. برای دنباله افزایشی از اعداد حقیقی بزرگتر از یک $\{\gamma_n\}_{n\geq 1}$، ناحیه نمونه‌گیری بی‌کران $D_n=\gamma_n D_.$ را در نظر بگیرید. با فرض آنکه دراین ناحیه، مشاهدات از یک میدان تصادفی مانای $\{Z(s):s\in\mathbb{Z}^d\}$ بر روی N_n موقعیت $\{s_1,\dots,s_{N_n}\}$ بصورت یک شبکه منظم باشند، آنگاه $N_n=\mathrm{Vol}(D_.)\cdot\gamma_n^d$ است، که برای سادگی با $N=N_n$ نمایش داده می‌شود. همچنین فرض کنید $\{\beta_n\}_{n\geq 1}$ یک دنباله از اعداد صحیح مثبت با فرض $\beta_n^{-1}+\lambda_n^{-1}\beta_n=o(1)$ باشد، که آنرا عامل مقیاس‌بندی برای بلوکها می‌نامیم. در روش خودگردانی بلوکی فضایی ناحیه نمونه‌گیری D_n با استفاده از مکعب‌هایی به حجم β_n^d بصورت $D_n(k)=\left[\beta_n(k+\boldsymbol{U})\right]\cap D_n$ افراز و بصورت $N=\beta_n^d\left|\boldsymbol{K}_n\right|$ کامل می‌شود، که در آن

$$\boldsymbol{K}_n=\left\{k\in\mathbb{Z}^d:\left[\beta_n(k+\boldsymbol{U})\right]\cap D_n\neq\phi\right\}$$

مجموعه اندیس بلوکهایی به حجم β_n^d و $\boldsymbol{U}=[0,1)^d$ مکعبی واحد در $\mathbb{R}^d$ است. بولمان و کونش (Buhlmann & Kunsch 1999) و ژو و لاهیری (Zhu & Lahiri 2001) روش MBB را برای برآورد مشخصات توزیع میانگین نمونه‌ای ارائه کردند. دراین روش ابتدا با استفاده از مجموعه بلوکهای مکعبی به حجم β_n^d از مجموعه اندیس $\boldsymbol{J}_n=\left\{j\in\mathbb{Z}^d:(j+\beta_n\boldsymbol{U})\subset D_n\right\}$ که متداخل هستند، تعداد $\left|\boldsymbol{K}_n\right|$ متغیر تصادفی مستقل $\left\{J_k:k\in\boldsymbol{K}_n\right\}$ با توزیع مشترک $P(J_1=j)=1/\left|\boldsymbol{J}_n\right|, j\in\boldsymbol{J}_n$ تولید می‌شود. سپس زیر نمونه MBB بصورت

$$\boldsymbol{Z}_n^*\left(D_n(k)\right)=\boldsymbol{Z}_n\left(\left[J_k+\beta_n\boldsymbol{U}\right]\cap\left[D_n(k)-k\beta_n+J_k\right]\right)$$

بدست می‌آید. حال نمونه MBB $\boldsymbol{Z}_n^*(D_n)$ از بهم پیوستن بلوکهای بازنمونه‌گیری شده $\boldsymbol{Z}_n^*\left(D_n(k)\right)$ تعیین می‌شود.

ایران‌پناه و محمدزاده (۱۳۸۴) روش SBB را براساس مجموعه بلوکهای مجزا با نقاط شروع از مجموعه اندیس $\boldsymbol{I}_n=\left\{i\in\mathbb{Z}^d:\beta_n(i+\boldsymbol{U})\subset D_n\right\}$ پیشنهاد کردند. برای بدست آوردن یک نمونه SBB، ابتدا $\left|\boldsymbol{K}_n\right|$ متغیر تصادفی مستقل $\left\{I_k:k\in\boldsymbol{K}_n\right\}$ با توزیع مشترک $P(I_1=i)=1/\left|\boldsymbol{I}_n\right|, i\in\boldsymbol{I}_n$ تولید می‌شود. سپس زیرنمونه SBB بصورت

$$\boldsymbol{Z}_n^*\left(D_n(k)\right)=\boldsymbol{Z}_n\left(\beta_n\left[I_k+\boldsymbol{U}\right]\cap\left[D_n(k)-k\beta_n+I_k\beta_n\right]\right)$$

بدست می‌آید. سرانجام نمونه SBB $\boldsymbol{Z}_n^*(D_n)$ از بهم پیوستن بلوکهای باز نمونه‌گیری شده $\boldsymbol{Z}_n^*\left(D_n(k)\right)$ تعیین می‌شود. حال اگر در این روش بجای مجموعه مشاهدات $\boldsymbol{Z}_n(D_n)=\left\{Z(s_1),\dots,Z(s_N)\right\}$ مجموعه $\boldsymbol{Y}_n(D_n)=\left\{\lambda_1 Z(s_1),\dots,\lambda_N Z(s_N)\right\}$، که در آن λ_i ها ضرایب کریگیدن $\hat{Z}_n(s_.)=\sum_{i=1}^{N}\lambda_i Z(s_i)$ هستند، را در نظر بگیریم و مجموعه خودگردانی بلوک مجزای $\boldsymbol{Y}_n^*(D_n)=\left\{\lambda_1^*Z^*(s_1),\dots,\lambda_N^*Z^*(s_N)\right\}$ را تعیین کنیم، نسخه SBB کریگیدن $\hat{Z}_n(s_.)$ بصورت $\hat{Z}_n^*(s_.)=\sum_{i=1}^{N}\lambda_i^*Z^*(s_i)$ بدست می‌آید. اریبی و واریانس خودگردانی $\hat{Z}_n^*(s_.)$ را می‌توان بترتیب بصورت $\mathrm{Bias}_*\left[\hat{Z}_n^*(s_.)\right]=E_*\left[\hat{Z}_n^*(s_.)\right]-\hat{Z}_n(s_.)$ و $\mathrm{Var}_*\left[\hat{Z}_n^*(s)\right]=E_*\left[\hat{Z}_n^*(s_.)-E_*\left(\hat{Z}_n^*(s_.)\right)\right]^2$ برآورد نمود، که در آن E_* و Var_* به ترتیب امید ریاضی و واریانس شرطی خودگردان به شرط $\boldsymbol{Z}_n$ است. اگر $\mathrm{Bias}_*\left[\hat{Z}_n^*(s_.)\right]$ و $\mathrm{Var}_*\left[\hat{Z}_n^*(s_.)\right]$ جوابهای صریحی نداشته باشند، آنها را می‌توان با استفاده از شبیه‌سازی مونت کارلو برآورد نمود.

خواص برآوردگر اندازه‌های دقت

فرض کنید برای میدان تصادفی مانای $\{Z(s):s\in\mathbb{Z}^d\}$، کریگیدن $\hat{Z}_n(s_.)=\sum_{i=1}^{N}\lambda_i Z(s_i)$ بعنوان پیشگوی مقدار $Z(s_.)$، براساس مشاهدات $\boldsymbol{Z}_n$ باشد. اگر $\hat{Z}_n^*(s_.)$ نسخه خودگردانی کریگیدن باشد، در اینصورت برآورد خودگردانی $\mathrm{Bias}\left[\hat{Z}_n(s_.)-Z(s_.)\right]$ که مقدار واقعی آن برابر صفر است، بصورت

$$\text{Bias}_*\left[\hat{Z}_n^*(s_.)-\hat{Z}_n(s_.)\right]=E_*\left[\hat{Z}_n^*(s_.)\right]-\hat{Z}_n(s_.)$$

تعیین می‌شود.

لم ۱: برآورد اریبی کریگیدن به روش SBB، برای یک میدان تصادفی مانای $\left\{Z(s):s\in\mathbb{Z}^d\right\}$ نااریب است.

برهان: مجموعهای $S_n(i;k)=\sum_{j\in B_n(i;k)}\lambda_j Z(j)$ و $S_n^*(k)=\sum_{j\in B_n(I_k;k)}\lambda_j^* Z^*(j)$ که در آنها $B_n(i;k)=D_n(k)-k\beta_n+i\beta_n$، iامین بلوک نوع k برای $k\in\boldsymbol{K}_n$ و $i\in\boldsymbol{I}_n$ باشد را در نظر بگیرید. بدلیل شانس حضور یکسان مشاهدات در بلوکهای مجزای $\left\{\beta_n(i+\boldsymbol{U}):i\in\boldsymbol{I}_n\right\}$ داریم:

$$\begin{aligned}E_*\left[\hat{Z}_n^*(s_.)\right]&=E_*\left[\sum_{k\in\boldsymbol{K}_n}S_n^*(k)\right]\\&=\left|\boldsymbol{K}_n\right|E_*\left[S_n^*(\cdot)\right]\\&=\left|\boldsymbol{K}_n\right|\left|\boldsymbol{I}_n\right|^{-1}\sum_{i\in\boldsymbol{I}_n}S_n(i;\cdot)\\&=\sum_{i=1}^{N}\lambda_i Z(s_i)=\hat{Z}_n(s_.)\end{aligned}\tag{۲}$$

بنابراین در روش SBB اریبی کریگیدن نااریب برآورد می‌گردد.

در روش SBB، اگر بجای مجموعه خودگردانی $\boldsymbol{Y}_n^*(D_n)$ در $\hat{Z}_n^*(s_.)=\sum_{i=1}^{N}\lambda_i^*Z^*(s_i)$ از نمونه خودگردانی $\boldsymbol{Z}_n^*(D_n)$ بصورت $\hat{Z}_n^*(s_.)=\sum_{i=1}^{N}\lambda_i Z^*(s_i)$ استفاده شود، خاصیت ذکر شده در لم ۱ محقق نخواهد شد. همچنین در روش MBB برخلاف روش SBB بدلیل شانس نابرابر مشاهدات برای حضور در بلوکهای متحرک $\left\{(j+\beta_n\boldsymbol{U}):j\in\boldsymbol{J}_n\right\}$ برآورد اریبی کریگیدن نااریب است، یعنی رابطه (۲) برقرار نمی‌باشد.

برآورد خودگردانی پارامتر $\sigma_n^2=N^{-1}\text{Var}\left(\hat{Z}_n(s_.)\right)$ را می‌توان بصورت

$$\hat{\sigma}_n^2=N^{-1}\text{Var}_*\left(\hat{Z}_n^*(s_.)\right)\tag{۳}$$

تعیین نمود. برای اثبات سازگاری $\hat{\sigma}_n^2$ در قضیه ۱ ابتدا لم ۲ معرفی می‌شود، که توسط دخان (Doukhan, 1994) اثبات شده 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 مقداری از $q\in\mathbb{N}$ 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 α−آمیختگی ($\alpha-\text{mixing}$) $Y(\cdot)$ است. آنگاه برای هر زیر مجموعه $A\in\mathbb{Z}^d$

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

است، که در آن ثابت $C(q,\delta)$ تنها به q، δ و d و ضریب α−آمیختگی $\alpha_Y(\cdot,\cdot)$ بستگی دارد، ولی به زیر مجموعه A بستگی ندارد.

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 α−آمیختگی $\alpha(a;b)\le Ca^{-\tau_1}b^{\tau_2}$ برای $a\ge1$ و $b\ge1$ و مقداری از $\delta\ge0$، $\tau_1\ge\frac{5d(6+\delta)}{\delta}$ و $0\le\tau_2\le\frac{\tau_1}{d}$ باشد. اگر $\beta_n^{-1}+\lambda_n^{-1}\beta_n=o(1)$ آنگاه

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

که در آن $\hat{\sigma}_n^2$ در (۳) تعریف شده و

$$\begin{aligned}\sigma_\infty^2&=\lim_{n\to\infty}N^{-1}\,\text{Var}\left(\hat{Z}_n(s_.)\right)\\&=\sum_{i\in\mathbb{Z}^d}\lambda_.\lambda_i E\left[Z(\cdot)-\mu\right]\left[Z(i)-\mu\right].\end{aligned}$$

برهان: بدون از دست دادن کلیت مسئله می‌توان میانگین μ را برابر صفر در نظر گرفت. با توجه به استقلال بلوکهای بازنمونه‌گیری شده داریم:

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

برای بررسی سازگاری $\hat{\sigma}_n^2$، کافی است نشان دهیم وقتی $n\to\infty$، آنگاه $E\left(\hat{Q}_{1n}\right)\to\sigma_\infty^2$، $E\left(\hat{Q}_{2n}\right)\to0$، $\text{Var}\left(\hat{Q}_{1n}\right)\to0$ و $\text{Var}\left(\hat{Q}_{2n}\right)\to0$.

الف)

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

ب) با توجه به لم ۱ داریم:

$$\begin{aligned} E\left(\hat{Q}_{۲n}\right) &= E\left[\beta_n^{-d}\left(E_*\left(S_n^*(\cdot)\right)\right)^{۲}\right] \\ &= \beta_n^{-d} E\left[N\left|\boldsymbol{K}_n\right|^{-۱}\hat{Z}_n(s_.)\right]^{۲} \\ &= \beta_n^{-d}\operatorname{Var}\left[\hat{Z}_n(s_.)\right] \\ &= O\left(\lambda_n^{-d}\beta_n^{d}\right) \end{aligned}$$

ج) با استفاده از لم ۲ برای $\delta = q = ۱$ داریم:

$$\begin{aligned} \operatorname{Var}\left(\hat{Q}_{۱n}\right) &= \operatorname{Var}\left[\beta_n^{-d} E_*\left(S_n^*(\cdot)\right)^{۲}\right] \\ &= \beta_n^{-۲d}\operatorname{Var}\left[\left|\boldsymbol{I}_n\right|^{-۱}\sum_{i\in \boldsymbol{I}_n}\left(S_n(i,\cdot)\right)^{۲}\right] \\ &= N^{-۲}E\left[\sum_{i\in \boldsymbol{I}_n}\left(S_n(i,\cdot)\right)^{۲} - E\left(S_n(i,\cdot)\right)^{۲}\right]^{۲} \\ &= N^{-۲}E\left[\sum_{i\in \boldsymbol{I}_n} V_n(i)\right]^{۲} \\ &\le N^{-۲}\left|\boldsymbol{I}_n\right|\max\left\{\left(E\left|V_n(i)\right|^{۳}\right)^{۲/۳} : i\in \boldsymbol{I}_n\right\} \\ &\quad \times\sum_{r=۰}^{\infty}(r+۱)^{d-۱}\alpha\left(r\beta_n;\beta_n^{d}\right)^{۱/۳} \\ &\le N^{-۲}\left|\boldsymbol{I}_n\right|\max\left\{E\left|S_n(i;\cdot)\right|^{۶} : i\in \boldsymbol{I}_n\right\}^{۲/۳} \\ &\quad \times\sum_{r=۰}^{\infty}(r+۱)^{d-۱}(r+۱)^{-\tau_۱/۳}\left(\beta_n^{\tau_۲ d-\tau_۱}\right)^{۱/۳} \\ &= O\left(N^{-۲}\left|\boldsymbol{I}_n\right|\left(\beta_n^{۳d}\right)^{۲/۳}\right) \\ &= O\left(\lambda_n^{-d}\beta_n^{d}\right) \end{aligned}$$

د)

$$\begin{aligned} \operatorname{Var}\left(\hat{Q}_{۲n}\right) &= \operatorname{Var}\left[\beta_n^{-d}\left(E_*\left(S_n^*(\cdot)\right)\right)^{۲}\right] \\ &= \beta_n^{-۲d}\operatorname{Var}\left[\beta_n^{۲d}\hat{Z}_n(s_.)\right]^{۲} \\ &= \beta_n^{۲d}\operatorname{Var}\left[\hat{Z}_n^{۲}(s_.)\right] \\ &= O\left(\lambda_n^{-۲d}\beta_n^{۲d}\right) \end{aligned}$$

با استفاده از مراحل (الف) و (ب) که نشان می‌دهند، برآوردگر $\hat{\sigma}_n^{۲}$ مجانباً ناریب است و مراحل (ج) و (د) برهان کامل می‌شود.

مقایسه دو روش SBB و MBB

در این بخش، دو روش SBB و MBB با استفاده از تکنیک شبیه‌سازی مونت کارلو برای داده‌های فضایی مورد مقایسه عددی قرار می‌گیرند.

فرض کنید $\left\{Z(s): s\in\mathbb{Z}^{۲}\right\}$ یک میدان تصادفی گاوسی مانای مرتبه دوم با میانگین صفر و همتغییرنگارهای کروی

$$\sigma(h;\theta)=\begin{cases} c_. + c & h=۰ \\ c\left[۱-\frac{۳}{۲}\left(\frac{h}{a}\right)+\frac{۱}{۲}\left(\frac{h}{a}\right)^{۳}\right] & ۰<h\le a \\ ۰ & h\ge a \end{cases}$$

و نمایی

$$\sigma(h;\theta)=\begin{cases} c_. + c & h=۰ \\ ce^{\frac{-h}{a}} & h\neq ۰ \end{cases}$$

باشد. با فرض آنکه $\theta=(c_.,c,a)=(۱,۲,۳)$ و داده‌های فضایی در یک شبکه منظم مستطیلی ۲۰×۳۰ $(N=۶۰۰)$ قرار دارند، با در نظر گرفتن فاصله اقلیدسی بین دو موقعیت، ماتریس کواریانس میدان تصادفی بصورت $\sum=\left(\sigma\left(s_i-s_j\right)\right)$ بدست می‌آید. اگر L تجزیه چولسکی (Choleski Decomposition) ماتریس $\sum$ بصورت $\sum=LL'$ باشد، یک نمونه تصادفی $\boldsymbol{Z}_n=\left(Z(s_۱),\ldots,Z(s_N)\right)$ از میدان تصادفی را می‌توان بصورت $\boldsymbol{Z}_n=L\boldsymbol{X}_n$ بدست آورد. که در آن $\boldsymbol{X}_n=(X_۱,\ldots,X_N)$ یک نمونه تصادقی مستقل از توزیع $N(۰,۱)$ است. مقادیر دقیق اریبی $T_۱=\left(\hat{Z}(s_.)-Z(s_.)\right)$ و واریانس کریگیدن $T_۲=\hat{Z}(s_.)$ در پنج موقعیت جدید $s_{.۱}=(۱/۵,۱/۵)$، $s_{.۲}=(۷/۵,۵/۵)$، $s_{.۳}=(۱۵/۵,۱۰/۵)$، $s_{.۴}=(۲۲/۵,۱۵/۵)$ و $s_{.۵}=(۲۹/۵,۱۹/۵)$ در ستون سوم جدول ۱ نشان داده شده است. همچنین مقادیر برآورد اریبی $T_۱$ و واریانس $T_۲$ به دو روش SBB و MBB در موقعیتهای مختلف $s_{.i}$ به ترتیب در دو ستون SBB و MBB جدول ۱ نشان داده شده است. برای انجام روش خودگردانی بلوکی فضایی، ناحیه نمونه‌گیری را بصورت $D_n=[-۱۰,۱۰)\times[-۱۵,۱۵)$ با ثابت مقیاس‌بندی $\gamma_n=۳۰$، اندازه بلوک $\beta_n=۵$ و مجموعه نمونه اولیه $D_.=\left[-\frac{۱}{۳},\frac{۱}{۳}\right)\times\left[-\frac{۱}{۲},\frac{۱}{۲}\right)\subset\left[-\frac{۱}{۲},\frac{۱}{۲}\right)^{۲}$ درنظر می‌گیریم. در این دو روش که با استفاده از مجموعه مشاهده شده $\boldsymbol{y}_n(D_n)=\left\{\lambda_۱ z(s_۱),\ldots,\lambda_N z(s_N)\right\}$ به حجم $N=۶۰۰$ که در یک شبکه منظم مستطیلی ۲۰×۳۰ از میدان تصادفی مورد نظر قرار گرفته‌اند، بلوکهایی به ابعاد ۵×۵ را به دو صورت مجزا و متحرک در نظر می‌گیریم. برای این منظور ناحیه نمونه‌گیری را به ۲۴ زیرناحیه یا بلوک بصورت

$$D_n(k)=[۵k_۱,۵k_۱+۵)\times[۵k_۲,۵k_۲+۵),\ k=(k_۱,k_۲)'\in\mathbb{Z}^{۲},$$
$$-۲\le k_۱<۲,\ -۳\le k_۲<۳$$

جدول ۱- مقادیر دقیق اریبی (و واریانس) آماره‌ها و برآورد آنها در موقعیتهای مختلف.

همتغییرنگار	موقعیت	مقدار واقعی	SBB	MBB
کروی	$s_{.1}$	۰/۰۰۰ (۱/۲۰۰۸)	۰/۰۰۹ (۱/۰۹۲۹)	۰/۷۴۱ (۰/۱۲۲۰)
	$s_{.2}$	۰/۰۰۰ (۱/۲۱۸۷)	۰/۰۰۹ (۰/۸۴۱۸)	۰/۴۰۰ (۱/۵۰۴۶)
	$s_{.3}$	۰/۰۰۰ (۱/۲۱۸۷)	۰/۰۰۸ (۰/۵۶۶۸)	۰/۳۹۶ (۱/۴۴۵۶)
	$s_{.4}$	۰/۰۰۰ (۱/۲۱۸۷)	۰/۰۰۸ (۰/۸۰۷۰)	۰/۳۷۵ (۱/۳۶۳۹)
	$s_{.5}$	۰/۰۰۰ (۱/۲۰۰۸)	۰/۰۰۹ (۱/۱۹۲۵)	۰/۷۶۸ (۰/۱۳۳۶)
نمایی	$s_{.1}$	۰/۰۰۰ (۱/۴۴۱۸)	۰/۰۱۱ (۱/۴۶۷۹)	۰/۸۴۲ (۰/۱۵۴۲)
	$s_{.2}$	۰/۰۰۰ (۱/۴۵۳۵)	۰/۰۰۸ (۰/۷۵۴۵)	۰/۳۸۸ (۱/۳۵۴۷)
	$s_{.3}$	۰/۰۰۰ (۱/۴۵۳۶)	۰/۰۰۸ (۰/۵۰۹۰)	۰/۴۳۳ (۱/۴۹۴۸)
	$s_{.4}$	۰/۰۰۰ (۱/۴۵۳۵)	۰/۰۰۸ (۰/۸۱۶۸)	۰/۴۰۴ (۱/۴۸۲۱)
	$s_{.5}$	۰/۰۰۰ (۱/۴۴۱۸)	۰/۰۱۰ (۱/۳۳۴۵)	۰/۷۹۳ (۰/۱۴۴۵)

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 بترتیب ۲۴ و ۴۱۶ است که از بین آنها به روش نمونه‌گیری تصادفی ساده با جایگذاری، ۲۴ بلوک به حجم ۲۵ را انتخاب و از به هم پیوستن آنها یک مجموعه خودگردانی $\boldsymbol{Y}_n^*(D_n)=\left\{\lambda_1^*Z^*(s_1),\ldots,\lambda_N^*Z^*(s_N)\right\}$ به حجم $N=600$ تولید می‌شود. اکنون می‌توان کریگیدن خودگردانی $\hat{Z}^*(s_.)=\sum_{i=1}^{N}\lambda_i^*Z^*(s_i)$ را محاسبه و با تکرار ۱۰۰۰ بار این الگوریتم، مقادیر $\hat{Z}_1^*(s_.),\ldots,\hat{Z}_{1000}^*(s_.)$ را بدست آورد. اریبی T_1 و واریانس T_2 با استفاده از کریگیدن خودگردان بصورت

$$\widehat{\text{Bias}}_*(T_1)=\bar{\hat{Z}}^*(s_.)-\hat{Z}(s_.)$$

$$\widehat{\text{Var}}_*(T_2)=\frac{1}{1000}\sum_{i=1}^{1000}\left[\hat{Z}_i^*(s_.)-\bar{\hat{Z}}^*(s_.)\right]^2$$

برآورد می‌شوند، که در آن $\bar{\hat{Z}}^*(s_.)=\frac{1}{1000}\sum_{i=1}^{1000}\hat{Z}_i^*(s_.)$ است.

همانطور که در جدول ۱ ملاحظه می‌شود، در همه موقعیتها روش SBB نسبت به روش MBB برای هر دو مدل همتغییرنگار کروی و نمایی مقدار اریبی T_1 را بسیار نزدیک به مقدار واقعی برآورد می‌نماید. مقدار واریانس برآورد شده T_2 در موقعیتهای $s_{.1}$ و $s_{.5}$ که مجاور مرزها هستند، روش SBB و در موقعیتهای مرکزی $s_{.2}$، $s_{.3}$ و $s_{.4}$ روشMBB به مقدار واقعی نزدیکتر می‌باشند. یعنی روش SBB واریانس را در موقعیتهای نزدیک مرزها با دقت بیشتری برآورد می‌کند، در حالیکه روش MBB واریانس را در موقعیتهای مرکزی بهتر برآورد می‌نماید. تولید نمونه تصادفی از میدان تصادفی مورد نظر به روش تجزیه چولسکی به همراه برآورد اریبی و واریانس پیشگوی فضایی کریگیدن به روش شبیه‌سازی مونت کارلو و همچنین الگوریتمهای SBB و MBB به همراه برآوردهای اریبی و واریانس آماره‌های مورد نظر با استفاده از نرم افزار S-PLUS برنامه نویسی شده‌اند.

بحث و نتیجه‌گیری

بصورت نظری و عددی نشان داده شد که برآورد اریبی به روش MBB همراه با خطاست، که علت آن شانس کمتر حضور مشاهدات مرزی ناحیه نمونه‌گیری در بلوکهای بازنمونه‌گیری شده نسبت به مشاهدات مرکزی می‌تواند باشد. روش SBB ارائه شده در این مقاله نه تنها اریبی کریگیدن را نااریب برآورد می‌کند، بلکه برآوردگر واریانس کریگیدن حاصل از این روش از خاصیت سازگاری برخوردار می‌باشد. مطالعه شبیه‌سازی در مقایسه نتایج حاصل از دو روش SBB و MBB بیانگر آن است که میزان دقت برآوردهای واریانس کریگیدن بستگی به موقعیتی دارد که در آن تخمین صورت می‌پذیرد، بطوریکه روش SBB در موقعیتهای مجاور مرزها واریانس کریگیدن را دقیقتر برآورد می‌کند، در حالیکه روش MBB در موقعیتهای مرکزی از کارایی بیشتری برخوردار است.

اگر موقعیتهای نمونه‌گیری بجای قرار گرفتن بر $\mathbb{Z}^d$ بصورت یک مشبکه منظم کامل، بر $\mathbb{R}^d$ بصورت نامنظم و غیرکامل باشند، در این

حالت هم الگوریتم SBB قابل اجراست. در این صورت اندازه نمونه SBB در تکرار بازنمونه‌گیری‌ها یکسان نخواهد بود، اما اگر N^* اندازه نمونه SBB دریک بازنمونه‌گیری باشد، مشابه لم ۱ می‌توان نشان داد $E_*\left(N^*\right)=N$. همچنین در روش خودگردانی بلوکی فضایی اگر ساختار همبستگی به عنوان مثال همتغییرنگار $\sigma\left(h;\theta\right)$ نامعلوم باشد، می‌توان از برآورد $\hat{\sigma}\left(h;\hat{\theta}\right)$ در الگوریتم استفاده نمود.

تشکر و قدردانی

نویسندگان از نظرات و پیشنهادات اصلاحی داوران محترم مجله که موجب بهبود این مقاله گردید و همچنین از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد کمال تشکر و قدردانی را دارند.

منابع:

ایران‌پناه، ن.، محمدزاده، م. ۱۳۸۴: روش بوت‌استرپ بلوک مجزا در آمار فضایی، نشریه علوم دانشگاه تربیت معلم. جلد ۵، **۴:** ۶۶۵-۶۵۳.

Buhlmann P., Kunsch H.R. 1999: Comments on “Prediction of Spatial Cumulative Distribution Functions Using Subsampling”. *Journal of the American Statistical Association.* **94:** 97-99.

Doukhan P. 1994: Mixing: Properties and Examples, vol. 85 of Lecture Notes in Statistics. Springer-Verlag. New York.

Efron B. 1979: Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics.* **7:** 1-26.

Hall P. 1985: Resampling a Coverage Pattern. *Stochastic Processes and Their Application.* **20:** 231-246.

Hall P. 1988: On Confidence Intervals for Spatial Parameters Estimated from Nonreplicated Data. *Biometrika.* **44:** 271-277.

Matheron G. 1963: Principles of Geostatistics. *Economic Geology.* **58:** 1246-1266.

Sjostedt-DeLuna S., Young S. 2003: The Bootstrap and Kriging Prediction Intervals. *Scandinavian Journal of Statistics.* **30:** 175-192.

Zhu J., Lahiri S.N. 2001: Weak Convergence of Blockwise Bootstrapped Empirical Processes for Stationary Random Fields with Statistical Applications, Preprint, Department of Statistics, Iowa State University, Ames, IA.