طرح نمونه‌گیری فضایی متعادل دو مرحله‌ای برای پیش‌گویی میدان‌های تصادفی

رامین خاورزاده، محسن محمدزاده[۱]

گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: تاریخ پذیرش:

چکیده: آمار فضایی علم تحلیل داده‌های وابسته فضایی است. در مطالعات محیطی گاهی با داده‌هایی وابسته سروکار داریم که همبستگی آن‌ها ناشی از موقعیت قرارگیری در یک فضای معین است. از طرفی در بررسی‌های نمونه‌ای فرض بر آن است که اعضای نمونه، از جامعه‌ای با واحدهای مستقل گرفته‌شده‌اند. این فرض در تمامی مراحل نمونه‌گیری، مدل‌سازی و تحلیل مورد استفاده قرار می‌گیرد. اما وقتی اعضای جامعه مورد مطالعه به‌نوعی وابسته باشند، تمامی مراحل آماری و حتی روش‌های نمونه‌گیری نیازمند بازنگری و لحاظ کردن ساختار همبستگی داده‌ها خواهند بود. در نمونه‌گیری کلاسیک برای نمونه‌گیری از یک متغیر، چنانچه متغیرهای کمکی وجود داشته باشند، برای ارتقاء کیفیت طرح نمونه‌گیری می‌توان از نمونه‌گیری متعادل استفاده کرد. در این مقاله نمونه‌گیری فضایی متعادل معرفی می‌شود که در آن از مؤلفه‌های موقعیت‌های فضایی به‌عنوان متغیرهای کمکی استفاده‌شده است. سپس در مطالعه‌ای شبیه-سازی نشان داده می‌شود کریگیدن بر اساس یک نمونه متعادل نسبت به روش‌های دیگر نمونه‌گیری متحمل خطای کمتری در پیش‌گویی فضایی می‌شود. در انتها نحوهٔ کاربست روش ارائه شده در کاربرد نشان داده می‌شود.

واژه‌های کلیدی: نمونه‌گیری فضایی، نمونه‌گیری متعادل دومرحله‌ای، نمونه‌گیری فضایی بهینه، روش مکعب.

رده‌بندی ریاضی (۲۰۱۰): ۶۲D۰۵ و ۶۲H۱۱.

۱- آدرس الکترونیکی نویسنده مسئول مقاله: mohsen_m@modares.ac.ir

۱- مقدمه

برخلاف روش‌های معمول آماری که برای تحلیل و نمونه‌گیری از استقلال مشاهدات به‌عنوان یک فرض اساسی استفاده می‌شود. در آمار فضایی، داده‌هایی مورد تحلیل قرار می‌گیرند که همبسته بوده و این همبستگی ناشی از موقعیت قرارگیری آن‌ها در فضای مورد مطالعه است که همبستگی فضایی[۱] نامیده می‌شود. مدل‌بندی داده‌های فضایی به‌طور معمول توسط میدان تصادفی[۲] $\{Z(s);s\in D\}$ انجام می‌شود، که در آن مجموعهٔ اندیس گذار D زیرمجموعه‌ای از فضای اقلیدسی dبعدی R^d، $d\geq 1$ است.

فرض کنید موقعیت‌های فضایی اعضای جامعه موردنظر به‌صورت $U=\{s_1,...,s_N\}$ نشان داده شوند و هدف پیش‌گویی مقدار میدان تصادفی در یک موقعیت دلخواه s_0 با استفاده از مشاهدات به‌دست آمده در نقاط نمونه‌گیری $\{s_1,...,s_n\}$ باشد. ازآنجاکه محل قرارگیری اعضای نمونه انتخاب شده، بر عملکرد پیشگو مؤثر است، انتخاب نمونه مناسب برای پیش‌گویی از اهداف نمونه‌گیری فضایی به‌حساب می‌آید.

طرح‌های نمونه‌گیری برای بررسی‌های نمونه‌ای مستقل به‌طور جامع در متون آماری موردبحث و بررسی قرارگرفته‌اند. کوکران و واتسون [۱]، سارندال و همکاران [۲]، تیله [۳] و چامبرز و کلارک [۴] شرح مبسوطی از مشخصه‌های طرح‌های نمونه‌گیری[۳] را ارائه کرده‌اند که شامل تعاریف روشنی از جامعه مورد مطالعه، واحدهای نمونه‌گیری، چارچوب نمونه‌گیری و چگونگی انتخاب نمونه است. طرح‌های مرتبط با پژوهش‌ها و سنجش‌های محیطی که در آن‌ها واحدهای جامعه لزوماً از یکدیگر مستقل نیستند، روش‌های نمونه‌گیری متداول را با چالش‌های جدیدی روبه‌رو کرده است. نمونه‌گیری از منابع محیطی به سیستم‌های پیچیده‌ای نیاز دارد. که در آن ممکن است با زمان یا مکان، حرکت یا تغییر کنند. به‌علاوه همواره ساختن چارچوب نمونه‌گیری قابل‌اعتماد، از جامعه هدف به‌سادگی میسر نیست. حتی ممکن است برخلاف روش‌های نمونه‌گیری معمول، چارچوب مشخصی از واحدهای جامعه موردنظر، قابل‌تعریف نبوده و تنها، واحدهای جامعه بر اساس مکان قرارگیری آن‌ها در محیط قابل‌تعریف باشند. در این صورت با توجه به وجود یک جامعه گسترده، تنها قادر به نمونه‌گیری بخش کوچکی از جامعه خواهیم بود. یکی از مهم‌ترین مسائلی که در چنین جوامعی باید به آن توجه شود، مشابهت واحدهای نزدیک‌تر ازلحاظ مکانی است. بنابراین لازم است مختصات موقعیت‌های فضایی واحدها در تحلیل نمونه‌ای

1- Spatial correlation
2- Random field
3- Sampling design

لحاظ شود. ازاین‌رو برای چنین جوامعی روش‌های جدیدی تحت عنوان روش‌های نمونه‌گیری فضایی معرفی‌شده‌اند.

در اولین مطالعات انجام‌شده در زمینهٔ نمونه‌گیری فضایی می‌توان به کار مک بارانتی و وبستر [۵] اشاره کرد. آن‌ها نشان دادند اگر هدف، پیش‌گویی و تعیین پهنه‌بندی فضایی متغیر مورد مطالعه با روش کریگیدن[1]، بدون در نظر گرفتن اثرات مرزی باشد، به کمک مشبکهٔ مثلثی[2] از موقعیت‌های موقعیت‌های نمونه‌ای، واریانس خطای پیش‌گویی تحت فرض‌های مانایی و همسانگردی[3] مینیمم می‌شود. در این میان طرح‌های نمونه‌گیری دیگری نیز با رویکرد طرح مبنا[4] معرفی‌شده‌اند که بین آن‌ها می‌توان به طرح نمونه‌گیری موزون باطرد واحدهای مجاور[5] [۶] و روش دنباله‌ای واحدهای ناحیه‌ای وابسته[6] [۷] و طرح‌های طبقه‌ای موزاییک‌بندی تصادفی تعمیم‌یافته[7] [۸] اشاره کرد. اما هر یک از این روش‌ها با محدودیت‌های مختلفی روبرو هستند. به‌عنوان‌مثال، روش "نمونه‌گیری موزون باطرد واحدهای مجاور" برای گرفتن نمونه از فضایی یک بعدی در نظر گرفته می‌شود. این روش تنها از قرار گرفتن واحدهای مجاور در نمونه ممانعت می‌کند. در روش "دنباله‌ای واحدهای ناحیه‌ای وابسته" نمونه‌ها به‌صورت دو بعدی در نظر گرفته می‌شوند، اما ازآنجایی‌که ناحیه‌های در نظر گرفته‌شده خصوصاً در قسمت‌های مرزی، تعداد واحدهای مجاور برابر ندارند، در عمل محاسبات با مشکل مواجه می‌شوند. به‌علاوه در روش "طرح‌های طبقه‌ای موزاییک‌بندی تصادفی تعمیم‌یافته" که تعمیمی از دو روش "طرح نمونه‌گیری طبقه‌ای موزاییک‌بندی" و "طرح طبقه‌ای موزاییک‌بندی تصادفی" است، تنها الگویی برای انتخاب مکان اعضای نمونه معرفی‌شده است که دارای بیشترین پوشش در ناحیهٔ موردنظر باشند. بیشترین تمرکز مطالعات در نمونه‌گیری فضایی[8] در سال‌های اخیر بر روی ساختارهای هندسی باهدف تولید یک مشبکه تصادفی و بهینه‌سازی نسبت به‌اندازه نمونه، واریانس و خودهمبستگی فضایی داده‌ها بوده است [۱۰-۹]. دیگرویجتر و همکاران [۱۱] و دوبی و همکاران [۱۲] روش‌های مختلف نمونه‌گیری کلاسیک را برای داده‌های فضایی مورد استفاده قرار دادند و نحوهٔ انجام، مزیت‌ها و معایب این طرح‌ها را به‌تفصیل بیان کردند. به‌هرحال در روش‌های معرفی‌شده همچنان رویکرد نمونه‌گیری همان اصول روش‌های کلاسیک در نمونه‌گیری فضایی بوده است. گرافستروم و تیله

1- Kriging
2- Triangulated lattice
3- Isotropy
4- Design based
5- Survey sampling avoiding contiguous units
6- Dependent area units sequential technique
7- Generalized random tessellation sampling
8- Spatial sampling

[۱۳] روش جدیدی تحت عنوان "نمونه‌گیری متعادل درجه دوم[1]" را معرفی کردند که در آن به‌جز در نظر گرفتن متعادل بودن متغیرهای کمکی، لحاظ کردن پراکندگی مشاهدات در فضای چارچوب نمونه‌گیری را نیز میسر ساختند.

روش‌های مختلف نمونه‌گیری فضایی به‌منظور اهداف و مسائل گوناگون مطرح‌شده‌اند. بخشی از این اهداف می‌تواند همان اهداف نمونه‌گیری کلاسیک، یعنی برآورد پارامترهای جامعه هدف، مانند میانگین، مقدار کل یا نسبت باشد. پارامترهایی از این قبیل را می‌توان با دقتی معمول از نمونه احتمالاتی به کمک رویکرد "طرح مبنا" برآورد کرد. از طرف دیگر هاینینگ [۱۴] نوعی نمونه‌گیری را برای بررسی ساختار همبستگی بر اساس تغییرنگار[2] یا هم‌تغییرنگار[3]، مطرح کرد که بخش دیگر اهداف نمونه‌گیری فضایی را تشکیل می‌دهد. هدف دیگر مطالعه، علاوه بر برآورد پارامترهای جامعه، می‌تواند پیش‌گویی فضایی در یک موقعیت فاقد مشاهده باشد. نمونه‌گیری فضایی با طراحی شبکه‌های سنجش نیز در ارتباط است. برای مثال برای تغییر محل ایستگاه‌های سنجش آلودگی هوا، باران‌سنجی، سنجش سطوح ازن در یک شهر، مکان‌یابی پارک‌ها، ایستگاه‌های اتوبوس، مقرهای آتش‌نشانی یا توزیع فضایی قرارگاه‌های پلیس برای امنیت بیشتر شهر، از مدلی آماری استفاده می‌شود. در واقع در چنین رویکرد استنباطی که بدان "مدل مبنا[4]" گفته می‌شود، هدف تعیین طرح بهینه نمونه‌گیری به‌منظور پیش‌گویی در موقعیت‌ها یا نواحی فاقد مشاهده، برآورد پارامترهای تابع کوواریانس فضایی، برآورد ضرایب رگرسیونی روند یا پوشش مناسب نمونه‌ای است. بنابراین اهداف مطالعه نقش به‌سزایی در تعیین طرح نمونه‌گیری مناسب دارند، که ممکن است در مسئله دیگر کاملاً متفاوت باشد. در این مقاله با استفاده از مفهوم نمونه متعادل در نمونه‌گیری کلاسیک و به‌کارگیری آن در نمونه‌گیری فضایی، نمونه‌ای بهینه به نام نمونه فضایی متعادل[5] معرفی می‌شود. برای دست‌یابی به این نمونه از تکنیک روش مکعبی[6] استفاده خواهد شد. آنگاه در مطالعه‌ای شبیه‌سازی روش نمونه‌گیری فضایی متعادل با نمونه‌گیری تصادفی ساده بر اساس ملاک میانگین توان دوم خطای کریگیدن مورد مقایسه قرار می‌گیرد. سپس مثالی کاربردی ارائه و در انتها به بحث و نتیجه‌گیری پرداخته می‌شود.

در این مقاله ابتدا نمونه‌گیری متعادل شده[7] موردبررسی قرار می‌گیرد سپس برای کاربردی کردن روش نمونه‌گیری متعادل، روش مکعبی توضیح داده می‌شود. سپس با ورود به آمار فضایی، کاربرد

1- 2nd order balanced sampling
2- Variogram
3- Covariogram
4- Model based
5- Balanced spatial sample
6- Cube method
7- Balanced sampling

روش نمونه‌گیری متعادل در آمار فضایی موردبررسی قرار می‌گیرد. برای این مهم با توضیحی مختصر درباره کریگیدن، این روش را برای پیش‌گویی فضایی[1] توضیح خواهیم داد. درنهایت با انجام یک مطالعه شبیه‌سازی روش‌های موجود نمونه‌گیری فضایی متعادل با یکدیگر مقایسه خواهند شد و با یک مثال کاربردی بهینگی روش معرفی‌شده موردبررسی قرار خواهد گرفت.

۲- نمونه‌گیری متعادل

یتس [۱۵] برای ارتقاء کیفیت طرح نمونه‌گیری نمونه متعادل را معرفی کرد. در نمونه‌گیری کلاسیک برای متغیر پاسخ Y، چنانچه متغیر کمکی Z وجود داشته باشد، نمونه‌ای از تحقق‌های متغیر پاسخ Y بر روی متغیر کمکی Z، متعادل نامیده می‌شود، هرگاه مقادیر Z به‌گونه‌ای انتخاب شوند که میانگین نمونه‌ای آن‌ها دقیقاً برابر مقدار واقعی میانگین جامعه Z باشد. رویال و هرسون [۱۶] شرط قوی‌تر انطباق گشتاورهای اول نمونه‌ای Z با گشتاورهای جامعه متناظرشان را مطرح کردند. بینش نهفته در پس این متعادل‌سازی، این است که، با انطباق گشتاورهای نمونه‌ای با گشتاورهای جامعه Z، تعادلی تقریبی روی Y ایجاد شود، تا نمونه متعادل انتخابی یک نمونه معرف از نمونه‌های مستخرج از جامعه هدف باشد. در واقع نمونه‌ای معرف نامیده می‌شود که منعکس‌کننده ویژگی‌های جامعه در تمام ابعاد موردنظر باشد به‌نحوی‌که نتایج حاصل از آن نمونه با نتایج حاصل از کل جامعه تقریباً همسان شوند [۱۷]. با این تعریف می‌توان با متعادل کردن نمونه، به یک نمونه معرف دست یافت.

فرض کنید جامعه موردنظر شامل N واحد آماری به‌صورت $U=\{۱,...,n\}$ و اطلاعات آن‌که معمولاً مقادیر تجمعی متغیرهای موردبررسی است، به‌صورت $t_y=(y_۱+...+y_N)=\sum_{k\in U} y_k$ نشان داده شوند. در نمونه‌گیری احتمالاتی، که انتخاب نمونه بر اساس احتمالات انجام می‌شود، احتمال انتخاب واحد u_k از جامعه به‌عنوان عضو نمونه عدد مشخص مثبت π_k است، که احتمال شمول[2] u_k نامیده می‌شود. بدیهی است مجموع احتمالات شمول برای کلیه واحدهای جامعه برابر با اندازه نمونه انتخابی است.

۲-۱- استخراج نمونه متعادل با روش مکعبی

روش مکعبی یکی از روش‌های مرسوم برای دست‌یابی به نمونه متعادل، با چند متغیر کمکی و با احتمالات شمول متفاوت است. این روش شامل دو مرحله فراز[3] و فرود[4] است. در مرحله فراز

1- Spatial prediction
2- Inclusion probability
3- Flight phase
4- Landing phase

قیدهای تعادل همواره برقرار هستند، به این صورت که مجموعه‌ای از نمونه‌ها در نظر گرفته می شود که همگی آن‌ها شرط تعادل را برقرار می‌کنند. اما در این نمونه‌ها اعضای جامعه به‌صورت کامل در نمونه قرار نمی‌گیرند. به‌عنوان‌مثال، چنانچه هدف انتخاب یک نمونه با اندازه ۲ از یک جامعه با ۴ عضو باشد. ممکن است یکی از نمونه‌ها در مرحله فراز به‌صورت نیمی از تمامی اعضای جامعه باشد، یعنی $\frac{1}{2}$ از واحد اول و $\frac{1}{2}$ از واحد دوم و $\frac{1}{2}$ از واحد سوم و $\frac{1}{2}$ از واحد چهارم که شرط تعادل را محقق می‌سازد. سپس با روشی کاملاً تصادفی یکی از اعضا به‌عنوان عضو نمونه که شرط تعادل را محقق می‌سازد، انتخاب می‌شود. در این مرحله چنانچه نمونه انتخاب‌شده به شکلی باشد که تمامی واحدهای قرارگرفته در آن به‌صورت کامل باشد، نمونه متعادل موردنظر انتخاب‌شده و نمونه‌گیری به پایان خواهد رسید. اما چنانچه واحدی در نمونه وجود داشته باشد که به‌صورت کامل در نمونه قرار نگرفته است، الگوریتم نمونه‌گیری وارد مرحله بعد می‌شود. در مرحله بعدی، یعنی مرحله فرود با توجه به‌احتمال شمول هر یک از اعضای جامعه اعضای نمونه‌ای که به‌صورت کسری در نمونه قرارگرفته‌اند به یکی از مقادیر یک (یعنی حضور در نمونه) یا صفر (یعنی عدم حضور این عضو در نمونه) تبدیل می‌شود. هدف این است که برای همه احتمال‌های شمول این کسرها به‌طور تصادفی به ۰ یا ۱ گرد شوند تا درنهایت به‌اندازه تعداد حجم نمونه، عضو کامل به دست آید. این مرحله، که مرحله فرود گفته می‌شود، لزوماً به برقراری کامل شرط تعادل منجر نمی‌شود بلکه تنها تعادلی نسبی روی متغیرهای کمکی ایجاد می‌کند.

۳- کریگیدن

به‌طور معمول با فرض مانایی میدان تصادفی، معلوم بودن ساختار همبستگی فضایی در قالب نیم تغییرنگار معلوم $\gamma(\cdot)$ و در نظر گرفتن تابع زیان درجه دوم، پیش‌گویی بهینه از مینیمم کردن میانگین توان دوم خطای پیشگو به دست آورده می‌شود. بهترین پیش‌گوی خطی $Z(s_0)$ تحت عنوان کریگیدن عادی به‌صورت ترکیب خطی

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) = \lambda' Z$$

تعریف می‌شود، که با مینیمم کردن میانگین توان‌های دوم خطای پیشگو، یعنی $E\,(Z\,(s_0) - \widehat{Z}\,(s_0))^2$، ضرایب آن به‌صورت

$$\widehat{\lambda'} = (\lambda + \underline{1}\frac{1 - \underline{1}'\Gamma^{-1}\underline{\gamma}}{\underline{1}'\Gamma^{-1}\underline{1}})\Gamma^{-1}$$

حاصل می‌شود، که در آن $\underline{\gamma}=(\gamma(s_0-s_1),...,\gamma(s_0-s_n))$، $\Gamma_{n\times n}=(\gamma(s_i-s_j))_{n\times n}$ و $\underline{1}$ یک بردار واحد $n\times 1$ است. واریانس این پیشگو که بیانگر میزان دقت آن است نیز به‌صورت

$$\hat{\sigma}_k^2(s_0)=\lambda'\Gamma^{-1}\underline{\gamma}+\frac{(1-\underline{1}'\Gamma^{-1}\underline{\gamma})^2}{\underline{1}'\Gamma^{-1}\underline{1}}$$

محاسبه می‌شود. درروش کریگیدن برای پیش‌گویی در یک موقعیت مشخص، به مشاهدات واقع در موقعیت‌های نزدیک‌تر وزن بیشتر و به مشاهدات دورتر وزن کم‌تری اختصاص داده می‌شود و وزن‌ها به‌گونه‌ای انتخاب می‌شوند که میانگین به توان دوم خطای پیشگو مینیمم شود [۱۸].

۴- طرح نمونه‌گیری فضایی متعادل دو مرحله‌ای

چنانچه موقعیت‌های آماری به‌صورت $U=\{s_1,...,s_N\}$ در نظر گرفته شوند و طرح نمونه‌گیری متعادل با رویکردی فضایی نگاه شود، شرط تعادل فضایی برای گشتاور اول به‌صورت $\sum_{s_k\in U}X_{s_k}=\sum_{s_k\in U}\frac{X_{s_k}C_{s_k}}{\pi_{s_k}}$ و برای گشتاور mام نیز به‌صورت $\sum_{s_k\in U}X_{s_k}^m=\sum_{s_k\in U}\frac{X_{s_k}^mC_{s_k}}{\pi_{s_k}}$ تعریف می‌شوند [۱۹].

اگر متغیرهای کمکی X_1 طول جغرافیایی، X_2 عرض جغرافیایی و X_3 ارتفاع واحدهای جامعه در نظر گرفته شوند، می‌توان با متعادل کردن بر اساس این متغیرهای کمکی به نمونه‌های بهینه متعادل رسید. وقتی متغیرهای کمکی موقعیت‌های فضایی در نظر گرفته شوند، نمونه‌ای متعادل فضایی است که گشتاورهای فضایی مختصات موقعیت‌های نمونه‌ای بر گشتاورهای فضایی جامعه منطبق باشند. گشتاورهای فضایی مرتبه اول و دوم مشابه با ماتریس کوواریانس است و نظم شکل ناحیه موردمطالعه یا نظم الگوی نقاط نمونه‌ای را اندازه می‌گیرند [۱۲]. نمونه‌گیری متعادل فضایی، عناصر نمونه‌گیری سیستماتیک و تصادفی ساده را با یکدیگر ترکیب می‌کند.

روش نمونه‌گیری متعادل متضمن نمونه‌ای است که به‌صورت کاملاً متعادل بر روی متغیرهای کمکی پخش‌شده باشد. حال چنانچه طول و عرض مختصات فضایی به‌عنوان متغیر کمکی در نظر گرفته شوند، نمونه حاصل، نمونه متعادل شده بر روی این متغیرهای کمکی است، اما این نمونه لزوماً در کل ناحیه موردبررسی پخش نشده است. برای رفع این مشکل، یک روش نمونه‌گیری متعادل دو مرحله‌ای پیشنهاد می‌کنیم.

در مرحله اول چارچوب فضایی با توجه به حجم نمونه طبقه‌بندی می‌شود به‌گونه‌ای که فضای نمونه‌ای به چند برابر (۵ تا ۱۰ برابر) حجم نمونه تقسیم‌بندی شود. طبیعتاً هرچه حجم نمونه بیشتر شود، طبقه‌بندی در مرحله اول بیشتر و امکان حصول نمونه متعادل اولیه بیشتر خواهد

بود. در مرحله دوم به تعداد اندازه نمونه طبقات با روش مکعبی انتخاب می‌شوند. این امر باعث می‌شود که طبقات نسبت به طول و عرض جغرافیایی متعادل باشند. سپس عضوی در هر طبقه انتخاب شده به‌عنوان نمونه انتخاب می‌شود که فاصله بیشتری را از طبقات انتخاب شده همسایه خود داشته باشد. به‌این‌ترتیب نمونه‌های نهایی، هم به لحاظ طول و عرض جغرافیایی متعادل هستند و هم بر روی فضای دوبعدی کاملاً پخش شده است. از طرف دیگر چون نمونه‌هایی که در ناحیه دو بعدی انتخاب می‌شوند در مرحله دوم بیشتر به سمت حاشیه‌ها رانده می‌شوند، نمونه‌های حاشیه‌ای تمایل بیشتری برای رفتن به حاشیه فضای دو بعدی خواهند داشت.

۵-مطالعه شبیه‌سازی

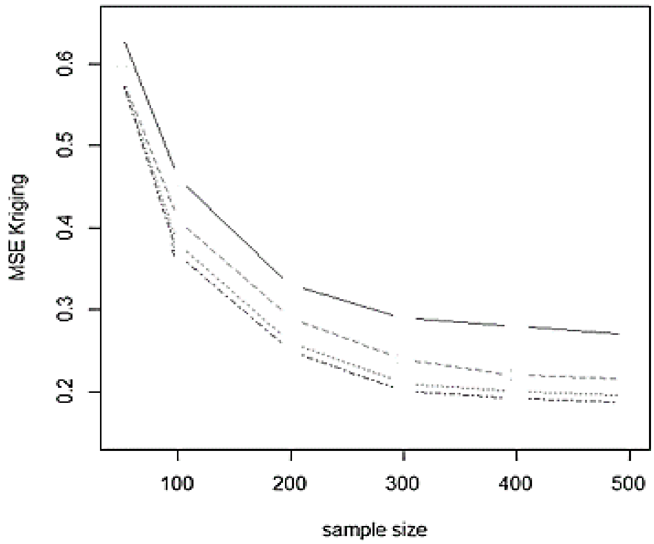
برای مقایسه نمونه‌گیری متعادل دو مرحله‌ای با روش‌های مرسوم مورد استفاده در نمونه‌گیری فضایی، یک ناحیه مشبکه‌ای ۱۰۰×۱۰۰ به‌وسیله نرم‌افزار R ساخته شد و با در نظر گرفتن نیم‌تغییرنگار کروی به‌صورت

$$\gamma(h)=\begin{cases} c_0 + c\left(\frac{3}{2}\frac{\|h\|}{a}-\frac{1}{2}\frac{\|h\|^3}{a^3}\right) & 0<\|h\|\le a,\ h\in R^2 \\ c_0 + c & \|h\|> a \end{cases}$$

مقدار پارامترهای مدل با دامنه ۲۰ و اثر قطعه‌ای ۰/۰۵ مقدار متغیر موردنظر در ۱۰۰۰۰ گره از میدان تصادفی گاوسی شبیه‌سازی شد. در مرحله بعد با فرض اینکه این گره‌ها مکان‌های قرارگیری متغیر موردبررسی در یک ناحیه ۱۰۰×۱۰۰ است، مقادیر دیگری به‌عنوان ارتفاع این نقاط به‌گونه‌ای از میدان تصادفی گاوسی شبیه‌سازی شد، که میزان همبستگی بین مقادیر موردبررسی و ارتفاع آن حداقل ۰/۴ باشد. سپس با روش نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه‌بندی‌شده، نمونه‌گیری متعادل توان دومی و نمونه‌گیری متعادل دو مرحله‌ای (که در این مقاله معرفی‌شده) تعداد ۵۰، ۱۰۰، ۲۰۰، ۳۰۰، ۴۰۰ و ۵۰۰ نمونه از اعضای جامعه گرفته شد و بر اساس نمونه‌های گرفته‌شده بقیه نقاط با روش کریگیدن پیش‌گویی شده است. نمونه‌گیری متعادل توان دومی روش نمونه‌گیری متعادلی است که از گشتاور دوم متغیرهای کمکی برای متعادل کردن نمونه استفاده می‌کند.

در مرحله بعد با در نظر گرفتن عرض جغرافیایی و توان دوم آن، طول جغرافیایی و توان دوم آن به همراه ارتفاع نقاط به‌عنوان متغیرهای کمکی، با روش مکعبی نمونه متعادل شده بر این ۵ متغیر کمکی ساخته‌شده است. ازآنجاکه روش مکعبی همواره نمونه‌ای با تعادل کامل نتیجه نمی‌دهد، برای هر نمونه معیاری به نام هزینه تعادل منظور شده است که هرچه مقدار آن به صفر نزدیک‌تر باشد بیانگر تعادل کامل نمونه است. بنابراین الگوریتم نمونه‌گیری مکعبی تا جایی تکرار

می‌شود که هزینه‌های تعادل بر متغیرهای کمکی به حداکثر ۰/۰۱ برسد. در این مرحله با در نظر گرفتن میانگین توان دوم خطای پیش‌گویی به‌عنوان ملاکی برای تشخیص نمونه بهینه، این معیار برای همه نمونه‌های گرفته‌شده محاسبه و مقادیر آن در شکل ۱ نشان داده‌شده است. همان‌طور که ملاحظه می‌شود، مقدار میانگین توان دوم خطای پیش‌گویی برای نمونه‌های متعادل دومرحله‌ای همواره کمتر از مقدار مشابه آن صدر نمونه‌های تصادفی و روش‌های دیگر بوده است.



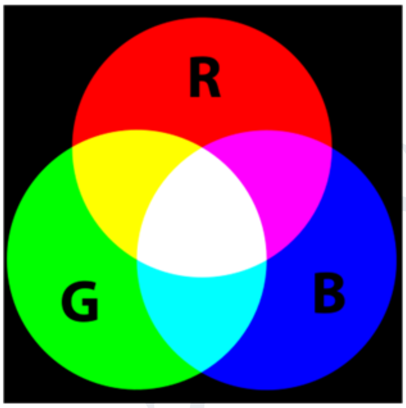
شکل (۱): نمودار مقادیر میانگین توان دوم خطای کریگیدن در مقابل اندازه‌ی نمونه با روش‌های نمونه‌گیری تصادفی ساده (خط ممتد)، طبقه‌بندی‌شده (خط‌چین)، GRTS (نقطه‌چین) و روش نمونه‌گیری متعادل فضایی (خط نقطه‌چین)

۶- تحلیل فضایی رنگ پیکسل‌های برج آزادی

برای ایجاد تصویر در تلویزیون، مانیتورها و صفحه‌های نمایش معمولاً مدل RGB به‌صورت کلی برای تصاویر به کار گرفته می‌شود. در این مدل، تمام رنگ‌ها از ترکیب سه رنگ ابتدایی قرمز (R)، سبز (G) و آبی (B) تشکیل می‌شود. با ترکیب این رنگ‌ها معمولاً رنگ‌های دیگر یا ثانویه[1] ایجاد می‌شوند (شکل ۲).

1- Secondary

از ترکیب سه رنگ ابتدایی RGB با رنگ سفید سیستم افزودنی[1] حاصل می‌شود. با توجه به اینکه صفحه‌نمایش این قبیل وسایل تیره هستند برای ایجاد تصویر باید به آن رنگ اضافه نمود. تبیین اینکه چرا به این نوع سیستم رنگ افزودنی می‌گویند از روش‌های مختلفی قابل توضیح است. رنگ سفید یا نور سفید داری طیف‌های مختلفی است که با یک منشور، آن طیف‌ها قابل‌تشخیص هستند. بنابراین می‌توان نتیجه گرفت که رنگ مشکی فاقد طیف نوری است. در یک زمینه مشکی برای ایجاد تصویر باید رنگ اضافه نمود.

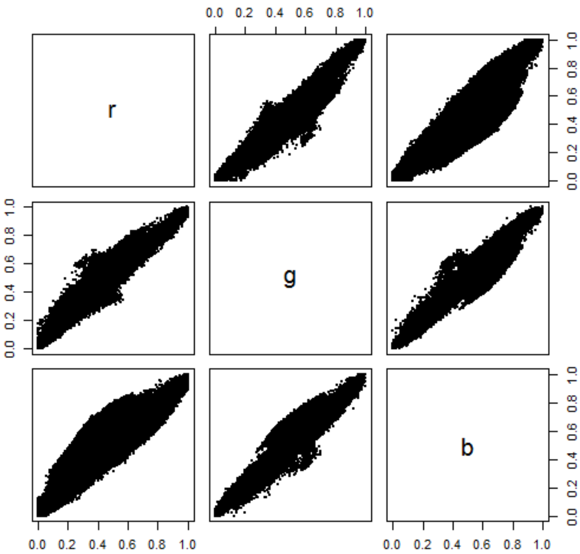


شکل (۲): نمایش رنگ‌ها به‌وسیله سه رنگ ابتدایی قرمز، آبی و سبز

به‌عنوان کاربردی از روش‌های نمونه‌گیری فضایی ابتدایی معرفی‌شده می‌توان به کاهش حجم عکس اشاره نمود. در این بخش عکس برج آزادی در شهر تهران (ایران) مورد مطالعه قرارگرفته است (شکل ۳). در ثبت این عکس با سیستم RGB، هر پیکسل شامل سه رنگ آبی، قرمز و سبز است. این عکس از ۲۲۶۲ × ۳۲۴۹ پیکسل تشکیل‌شده است.

1- Additive

شکل (۳): نمایش تصویر برج آزادی با تفکیک رنگ‌های آبی، قرمز، سبز مدل RGB



شکل (۴): نمودار پراکنش سه رنگ آبی (b) ،قرمز (r) و سبز (g) نسبت به یکدیگر

وجود همبستگی فضایی در این سه تفکیک عکس کاملاً مشهود است زیرا که احتمال وجود پیکسل‌های روشن در کنار پیکسل‌های روشن بیشتر و احتمال پیکسل‌های تیره در کنار پیکسل‌های تیره نیز بیشتر است. یا به‌عبارت‌دیگر پیکسل‌های نزدیک به هم بیشتر به یکدیگر نزدیک هستند تا پیکسل‌های که از یکدیگر فاصله بیشتری دارند. اما جدای از همبستگی فضایی، این سه رنگ ازلحاظ مکانی، همبستگی متقابل بین سه رنگ وجود دارد. در شکل ۴ نمودار پراکنش سه رنگ با یکدیگر مقایسه شده است. کاملاً واضح است که ارتباط بین رنگ‌ها با یکدیگر به‌صورت مستقیم و خطی است. ازاین‌رو می‌توان برای پیش‌گویی فضایی در هر پیکسل از هر سه رنگ در هر نقطه نیز استفاده کرد.

چنانچه رنگ‌هایی که در عکس برج آزادی به‌کاررفته است را به‌عنوان یک چارچوب فضایی با سه متغیر در نظر بگیریم، با استفاده از سه روش عنوان‌شده در این مقاله، به ترتیب نمونه‌هایی با حجم‌های ۲۰، ۴۰، ۷۰ و ۱۰۰ در نظر گرفته و برای هرکدام هزار نمونه منظور شده است. در مرحله بعدی با برازش توزیع فضایی مانند بخش شبیه‌سازی در همین مقاله، به داده‌های نمونه‌گیری شده مقادیر سه رنگ را در نقاط دیگر فضا، پیش‌گویی کرده‌ایم. جدول زیر نتایج معیارهای ریشه توان دوم خطای پیش‌گویی فضایی در سه روش و با اندازه‌های نمونه‌ای متفاوت نشان داده‌شده است.

جدول (۱): ریشه توان دوم خطای پیش‌گویی‌های حاصل از سه روش نمونه‌گیری فضایی

روش نمونه‌گیری	حجم نمونه			
	۲۰۰۰۰	۴۰۰۰۰	۷۰۰۰۰	۱۰۰۰۰۰
تصادفی ساده	۰/۵۷۳	۰/۴۲۶	۰/۳۳۴	۰/۲۶۵
متعادل شده مکعبی	۰/۴۵۳	۰/۳۵۴	۰/۲۸۷	۰/۲۸۵
متعادل شده دو مرحله‌ای	۰/۴۳۵	۰/۳۴۸	۰/۲۷۱	۰/۲۳۳

در جدول ۱ مقدار RMSE برای پیش‌گویی‌های حاصل از سه روش نمونه‌گیری تصادفی ساده، نمونه‌گیری متعادل و نمونه‌گیری دومرحله‌ای ارائه شده است. با توجه به‌اندازه نمونه‌های متفاوت ملاحظه می‌شود که روش دومرحله‌ای توانسته خطای کمتری نسبت به روش‌های پیشین ایجاد کند.

۷- بحث و نتیجه‌گیری

در این مقاله روش نمونه‌گیری متعادل فضایی دو مرحله‌ای پیشنهادشده و با استفاده از روش‌های پیش‌بینی آمار فضایی، سه روش نمونه‌گیری متداول برای نمونه‌گیری فضایی در مطالعه‌ای

شبیه‌سازی مورد مقایسه قرار گرفت. بر اساس این مقایسه روش نمونه‌گیری فضایی در تمامی حجم‌های نمونه دارای بیشترین خطای پیش‌گویی بود اما دو روش نمونه‌گیری فضایی دومرحله‌ای و روش نمونه‌گیری متعادل درجه دوم تا حدی به یکدیگر نزدیک بودند اما با اختلاف جزئی می توان گفت که روش نمونه‌گیری فضایی دومرحله‌ای معرفی‌شده در این مقاله توانست نتایج بهتری را در پیش‌گویی حاصل نماید. دلیل این موضوع را نیز می‌توان به پخش کامل نمونه در فضای مطالعه و همچنین در نظر گرفتن احتمال شمول بیشتر برای واحدهای حاشیه‌ای دانست. روش نمونه‌گیری فضایی دومرحله‌ای تا حدی احتمال انتخاب نمونه‌های حاشیه‌ای را بیشتر از نقاط داخلی در نظر می‌گیرد. از طرفی در پیش‌بینی فضایی بیشترین خطا در حاشیه فضاهای پیش‌بینی رخ می‌دهد و همین امر باعث بهتر عمل کردن این روش شده است. علاقه‌مندان می‌توانند برای دسترسی به کدهای شبیه‌سازی و نکات مرتبط مقاله با نویسنده مسئول مقاله مکاتبه فرمایند.

تقدیر و تشکر

نویسندگان از داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله شد کمال قدردانی را دارند.

منابع

[1] Cochran, W. G. and Watson, D. J. (1936). An experimental on observer's basis in the selection of shoot heights, *Empire Journal of Experimental Agriculture*, **4**, 69-76.

[2] Sarndal, C. E. and Swensson, B. and Wretman, J. (1992). Model Assisted, *Survey sampling*, Springer, New York.

[3] Tillé, Y. (2006). *Sampling algorithms*, Spinger-Verlag, New York.

[4] Chambers, R. and Clark, R. (2012), *An introduction to model-based survey sampling with applications*, Oxford University Press.

[5] McBratney, A. B. and Webster, R. (1981). Detection of ridge and furrow pattern by spectral analysis of crop yield, *International Statistical Review*, **49**, 45-52.

[6] Hedayat, A. S. and Majumdar, D. (1995). Generating desirable sampling plans by the technique of trade-off in experimental design, *Journal of Statistical Planning and Inference*, **44**, 237-247.

[7] Arabia, G. (1993). Use of GIS in spatial statistical surveys, *International Statistical Review*, **61**, 339-359.

[8] Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources, *Journal of the American Statistical Association*, **99**, 262-15.

[9] Dessard, H. and Bar-Hen, A. (2005). Experimental design for spatial sampling applied to the study of tropical forest regeneration, *Canadian Journal of Forest Research*, **35**, 1149-1155.

[10] Salehi, M. (2004). Optimal sampling design under a spatial correlation model, *Journal of Statistical Planning and Inference*, **118**, 9-18.

[11] De Gruijter, J., Brus, D., Bierkens, M. and Knotters, M. (2006). *Sampling for natural resource monitoring*, Springer, Berlin.

[12] Dobbie, M. J., Henderson, B. L. and Stevens, D. L. (2008). Sparse sampling: spatial design for monitoring stream networks, *Statistics Surveys*, **2**, 113-153.

[13] Grafström, A. and Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals, *Environmetrics*, **24**, 120-131.

[14] Haining, R. P. (2003). *Spatial data analysis: Theory and practice*, Cambridge University Press, Cambridge.

[15] Yates, F. (1935). Some examples of biased sampling, *Annals of Eugenics*, **6**, 202- 213.

[16] Royall, R. M. and Herson, J. (1973). Robust estimation in finite populations, *Journal of American Statistical Association*, **68**, 880-889.

[17] Appel, B. R., Tokiwa, Y. and Haik, M. (1967). Sampling of nitrates in ambient air, *Atmospheric Environment*, **15**, 283-289.

[18] Mateu, J. and Muller, W. (2012). Spatio-temporal design, advances in efficient data acquisition, John Wiley & Sons, Chichester, UK. ISBN: 978-0-470-97429-2.

[۱۹] محمدزاده، م. (۱۳۹۱). *آمار فضایی و کاربردهای آن*، مرکز نشر آثار علمی، دانشگاه تربیت مدرس، تهران، ایران.

Two-Step Spatial Balanced Sampling for Random Fields Prediction

Ramin Khavarzadeh and Mohsen Mohammadzadeh

Department of Statistics, Tarbiat Modares University, Tehran, Iran

Abstract

Spatial statistics is the science of analyzing spatial data. In environmental studies, we sometimes deal with dependent data which their correlation is due to the position in a given space. On the other hand, in sample surveys, it is assumed that sample are taken from a community of independent units. This assumption is used in all steps of sampling, modeling and analysis. But when the members of the studied community are in some way dependent, all statistical steps and even sampling methods will need to be reviewed and incorporate the structure of data correlation. In classical sampling for sampling from a variable, in presence of covariates, balanced sampling can be used to improve the quality of the sampling plan. In this paper, a spatial balanced sampling is introduced in which the components of spatial positions are used as covariates. Then, by a simulation study, it is shown that Kriging based on a balanced sample produces less error in spatial prediction than other sampling methods. At the end, the application of this method is shown.

Keywords: Spatial sampling, Two-step spatial balanced sampling, Optimal spatial sampling, Cube method.

Mathematics Subject Classification (2010): 62D05, 62H11