

تابع کوواریانس تعمیم یافته برای کریگیدن عام

محسن محمدزاده^۱

چکیده

روش کریگیدن که برای برآورد داده‌های مکانی بکار برد می‌شد، وابسته به تابع کوواریانس می‌باشد. معمولاً این تابع که تعیین کننده ساختار همبستگی مشاهدات مکانی است، نامعلوم و برآورده آن امری ضروری می‌باشد. اغلب در کریگیدن عادی که میانگین مشاهدات ثابت فرض می‌شود تابعی به کوواریانس تجربی داده‌ها برآش داده می‌شود. همین روش نیز برای کریگیدن عام، با استفاده از داده‌هایی که روند آنها حذف گردیده، بکار برد می‌شود. اما بدلیل آنکه تابع کوواریانس داده‌های روند زدوده و داده‌های اصلی متفاوت هستند این روش مشکل آفرین می‌باشد. در این مقاله نشان داده می‌شود داده‌های روند زدوده و داده‌های اصلی دارای تابع کوواریانس تعمیم یافته یکسانی هستند و می‌توان آنرا بر اساس کوواریانس تجربی داده‌های روند زدوده برآورد نمود. چون کریگیدن حاصل از تابع کوواریانس معمولی و تعمیم یافته معادل یکدیگرند برآورده کوواریانس تعمیم یافته را می‌توان برای محاسبه کریگیدن بکار برد.

کلمات کلیدی : داده‌های فضایی، کریگیدن عام و کوواریانس تعمیم یافته.

۱ مقدمه

$$\text{داده‌های فضایی}^2 \text{ را می‌توان مقادیر یک میدان تصادفی$$

$Z(t) : t \in D \subseteq R^d$ در نظر گرفت که در آن t یک موقعیت در

مجموعه اندیس گذار D ، زیر مجموعه‌ای از فضای اقلیدسی است. همچنین برای دو موقعیت t_1 و t_2 در D تابع کوواریانس $Z(t_1)$ و $Z(t_2)$ که بصورت:

$$\sigma(t_1, t_2) = \text{Cov}(Z(t_1), Z(t_2)) \quad (3)$$

$d \geq 1$ است و به ازاء هر t ، $Z(t)$ یک متغیر تصادفی با

میانگین و واریانس

$$\mu(t) = E[Z(t)] \quad (1)$$

^۱محسن محمدزاده، گروه آمار، دانشگاه تربیت مدرس
^۲Spatial Data

ارائه خواهد شد، برآورده می‌شود. چون معمولاً تابع کوواریانس نامعلوم است، برآورده آن برای تجزیه و تحلیل داده‌های فضایی، بخصوص کریگیدن، امری ضروری می‌باشد. در روش کریگیدن عادی اغلب با استفاده از روش‌های گرافیکی، تابعی به کوواریانس تجربی

$$\hat{o}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(t_i) - \tilde{Z})(Z(t_j) - \tilde{Z}) \quad (5)$$

برازش می‌شود. در عبارت (5) مجموع روی تمام اعضای مجزای مجموعه $\{(i, j) : t_i - t_j = h\}$ گرفته می‌شود و Nh نشان دهنده فاصلهٔ دو موقعیت، \tilde{Z} یک برآورده میانگین Z و $|N(h)|$ تعداد اعضای $N(h)$ می‌باشد.

دلفینر (۱۹۷۶) نشان داد وجود روند در داده‌ها کریگیدن عام را با مشکل تعیین تابع کوواریانس مواجه می‌نماید. بعلاوهٔ کرسی و گروندونا (۱۹۹۲) برای این حالت تحت شرایط خاص جمع پذیری تابع کوواریانس نشان دادند برآورده تابع کوواریانس دارای اربیتی از مرتبهٔ $O(1/n)$ است. لذا اغلب برای برآورده تابع کوواریانس در کریگیدن عام، ابتدا روند را از داده‌ها حذف نموده، سپس همان رویه فوق بر داده‌های روند زدوده اعمال می‌شود. اما آرمسترانگ (۱۹۸۴) نشان داد کوواریانس تجربی داده‌های بدون روند شده و داده‌های اصلی متفاوت هستند و کاملاً بستگی به روش روندزدایی داده‌ها دارد. بنابراین این رویه نیز موجب بروز اربیتی در برآورده کوواریانس می‌گردد.

در بخش ۳ نشان داده می‌شود داده‌های اصلی و داده‌هایی که روند آنها حذف می‌گردد دارای تابع کوواریانس تعمیم یافته^۹ (GCF) یکسانی هستند و کوواریانس تجربی داده‌هایی که روند آنها حذف شده را فقط می‌توان برای برآورده تابع کوواریانس تعمیم یافته بکار برد و قابل استفاده برای برآورده تابع

تعريف می‌شود، نشان دهنده ساختار همبستگی داده‌های فضایی می‌باشد.

به ازاء هر t ، میدان تصادفی $Z(t)$ را می‌توان بصورت

$$Z(t) = \mu(t) + \sigma(t) \quad (4)$$

تجزیه نمود که در آن $\sigma(t)$ یک میدان تصادفی با میانگین صفر است و $\mu(t)$ را قسمت قطعی^۲ و یا روند^۳ و $\sigma(t)$ را قسمت تصادفی مدل (۴) می‌نامند. چنانچه میانگین و تابع کوواریانس میدان تصادفی بصورت:

$$\mu(t) = \sigma(t_1, t_2) = \sigma(t_1 - t_2) \quad \text{و}$$

باشد، یعنی میانگین مقداری ثابت و کوواریانس تابعی از فاصلهٔ دو موقعیت t_1 و t_2 باشد، آنرا میدان تصادفی مانای مرتبهٔ دوم^۵ می‌نامند.

فرض کنید مشاهدات $Z(t_1), \dots, Z(t_n)$ از میدان تصادفی $Z(\cdot)$ در n موقعیت فضایی (t_1, \dots, t_n) در اختیار باشد. یکی از متداولترین روش‌های برآورده فضایی برای مقدار میدان تصادفی در یک موقعیت مشخص t ، بهترین برآورده خطی ناریب (BLUP) براساس مشاهدات است که پس از کریگ (۱۹۵۱) مهندس معدن آفریقای جنوبی، توسط ماترون (۱۹۶۳) کریگیدن^۶ نامیده شده است.

کریگیدن شامل دو نوع متداول کریگیدن عادی^۷ و کریگیدن عام^۸ است که ماترون (۱۹۶۹ و ۱۹۷۳) و کرسی (۱۹۹۳) آنها را مورد بررسی مفصل قرار داده‌اند. در روش کریگیدن عادی، فرض می‌شود میدان تصادفی مانای مرتبهٔ دوم با میانگین ثابت است. اما در روش کریگیدن عام که تعمیم کریگیدن عادی است، فرض می‌شود میانگین میدان تصادفی ترکیب خطی توابعی از قبل تعیین شده است.

در هر دو روش کریگیدن، با استفاده از تابع کوواریانس $\sigma(h)$ و مشاهدات $Z(t)$ ، مقدار

Deterministic ^۳
trend ^۴
Second Order Stationary ^۵
Kriging ^۶
Ordinary Kriging ^۷
Universal Kriging ^۸
Generalized Covariance Function ^۹

با عناصر زیر است.

$$x_{ij} = \phi_{j-1}(t_i) \quad i = 1, \dots, n; \quad j = 1, \dots, p+1$$

کریگیدن در موقعیت t عبارتست از انتخاب یک تخمین‌گر خطی بصورت:

$$\hat{Z}(t_0) = \lambda^T Z \quad (8)$$

که بطور یکنواخت ناارب باشد، یعنی

$$E(\hat{Z}(t_0)) = E(Z(t_0)); \quad \beta \in R^{p+1} \quad (9)$$

باشد و میانگین مجدور خطاهای

$$\sigma_e^2 = E[(\hat{Z}(t_0) - Z(t_0))^2]$$

را نسبت به λ کمینه نماید. شرط بطور یکنواخت ناارب (9) معادل آنست که $x_0^T \lambda^T X = x_0^T \lambda$ باشد که در آن $\lambda = (\phi_0(t_0), \dots, \phi_n(t_0))^T$ است.

کریگیدن در نقطه t_0 بصورت (8) است که مقدار بهینه λ از رابطه:

$$\lambda = \Sigma^{-1} [\sigma_0 + X(X^T \Sigma^{-1} X)^{-1} (x_0 - X^T \Sigma^{-1} \sigma_0)]$$

که در آن $(\sigma_0, \dots, \sigma_n) = (\sigma(t_0 - t_1), \dots, \sigma(t_0 - t_n))$ است، محاسبه می‌گردد. بعلاوه مقدار کمینه میانگین مجدور خطای برآورد برابر

$$\begin{aligned} \sigma_e^2(t_0) &= \sigma_0^2 - \sigma_0^T \Sigma^{-1} \sigma_0 + (x_0 - X^T \Sigma^{-1} \sigma_0)^T \\ &\quad \times (X^T \Sigma^{-1} X)^{-1} (x_0 - X^T \Sigma^{-1} \sigma_0) \end{aligned}$$

می‌باشد. بدیهی است اگر $p = 1$ و $\phi_0(t) = \phi$ اختیار شود، کریگیدن عام به کریگیدن عادی تبدیل می‌گردد.

کوواریانس معمولی نیستند. از طرفی ماترون (1971) نشان داده است برآوردهای حاصل از تمام توابع کوواریانس تعمیم یافته یکسان هستند. همچنین کریستنسن (1990) نشان داد برآوردهای حاصل از کریگیدن با استفاده ازتابع کوواریانس معمولی و GCF معادل یکدیگردند. بنابراین کریگیدن عام را می‌توان بوسیله برآورد GCF که بر اساس داده‌های روندزدوده بدست آورده می‌شود، انجام داد.

۲ کریگیدن

ماترون (1971) کریگیدن عادی را برای برآورد فضایی میدان تصادفی در یک نقطه مشخص t_0 بر اساس بردار مشاهدات Z و با فرض آنکه میانگین یا روند در مدل (4) مقدار ثابت $\mu = \mu(t)$ و $\delta(t)$ یک میدان تصادفی مانای مرتبه دوم باشد، توسط ترکیب خطی $\hat{Z}(t_0) = \sum_{i=1}^n \lambda_i Z(t_i)$ مشروط بر آنکه $\sum_i \lambda_i = 1$ باشد معرفی کرد، که شرط $\sum_i \lambda_i = 1$ بودن برآوردهای تضمن می‌نماید و ضرایب برآوردهای نیز به گونه‌ای انتخاب می‌شوند که میانگین مجدور خطاهای کمینه گردد.

در روش کریگیدن عام که توسط ماترون (1969) ارایه گردیده است فرض می‌شود روند در مدل (4) یک ترکیب خطی از توابعی از قبل تعیین شده مانند (ϕ_0, \dots, ϕ_p) با ضرایب نامعلوم بصورت:

$$\mu(t) = \sum_{j=0}^p \beta_j \phi_j(t) \quad (6)$$

است، که در آن توابع $\{\phi_0(t), \dots, \phi_p(t)\}$ با توجه به ماهیت روند تعیین می‌شوند. در این حالت مدل (4) را می‌توان به فرم ماتریسی

$$Z = X\beta + \delta \quad (7)$$

نوشت که در آن توابع $(\delta(t_1), \dots, \delta(t_n)) = \delta$ و $n \times (p+1)$ یک ماتریس $X = (\beta_0, \dots, \beta_p)^T \in R^{p+1}$

می باشد. در کریگیدن عام تعداد قیود نااریبی برابر تعداد ضرایب نامعلوم روند خطی، یعنی p در عبارت (۶) است. هرچقدر تعداد قیود نااریبی بیشتر باشد، تعداد توابع کوواریانسی که کاملاً یکسان عمل می کنند بیشتر می شود. عنوان مثال برای کریگیدن عام با روند خطی، دو تابع کوواریانس

$$\sigma(h) = \theta_1 \exp(-h/\theta_2) \quad (13)$$

$$\sigma(h) = \theta_1 \exp(-h/\theta_2) + a_1 + a_2 h^2 \quad (14)$$

که در آن a_1 و a_2 ثابت های دلخواه هستند، همانند می باشند.

در این حالت تابع کوواریانس را می توان بصورت:

$$\sigma(h) = C(h) + l(h) \quad (15)$$

نشان داد که در آن $C(h)$ قسمت اصلی یا GCF و $l(h)$ قسمت افزونه نامیده می شود. عبارت (۵۱) که GCF را عنوان قسمت اصلی تابع کوواریانس معرفی می نماید، در محاسبه کریگیدن عام با روندی شامل ضرایب نامعلوم و همچنین ضرایب روش های برآورد تابع کوواریانس مربوطه بسیار مفید می باشد. بعلاوه افزایش تعداد قیود نااریبی، آزادی بیشتری برای انتخاب یک تابع کوواریانس فراهم می سازد. بطور کلی، با معرفی جملات بیشتر در روند یا قسمت قطعی مدل (۴)، عمل جملات بیشتری از قسمت اصلی تابع کوواریانس به قسمت افزونه انتقال داده می شود، زیرا با بیان بیشتر تغییرات از طریق یک تابع قطعی، اهمیت تابع کوواریانس (که قسمت تصادفی مدل را تشریح می کند) کاهش می یابد. در حد، وقتی تمام تغییرات مدل را از طریق یک تابع قطعی بیان نماییم، GCF فقط به یک اثر قطعه $^{\circ}$ ساده تبدیل می گردد.

اکنون نشان داده می شود داده های اصلی و داده های بدون روند دارای تابع کوواریانس متفاوتی هستند ولی GCF آنها یکسان است. فرض کنیم Z ماتریس کوواریانس داده های اصلی، یعنی Z ، با عناصر

$$\sigma_{ij} = \sigma(t_i - t_j), \quad i, j = 1, \dots, n$$

۳ تابع کوواریانس تعمیم یافته

در روش کریگیدن عادی، معلوم بودن تابع کوواریانس تا یک مقدار ثابت کفایت می کند. عنوان مثال، کریگیدن برای دو تابع کوواریانس

$$\sigma(h) = \theta_1 \exp(-h/\theta_2) \quad (10)$$

$$\sigma(h) = \theta_1 \exp(-h/\theta_2) + a \quad (11)$$

که در آنها θ_1 و θ_2 پارامتر و a یک مقدار ثابت دلخواه هستند، نتایج یکسانی را تولید می کند. علت عدم تأثیر مقدار ثابت a وجود قید نااریبی است، زیرا اضافه کردن یک مقدار ثابت به تابع کوواریانس معادل آنست که به میدان تصادفی (Z) یک میدان تصادفی با میانگین صفر واریانس ثابت a اضافه نماییم. چون اضافه کردن یک مقدار ثابت به میدان تصادفی (Z) در خطای برآورد بی تأثیر است، a هیچ تغییری در کریگیدن ایجاد نمی کند. پس برای کریگیدن عادی، تمام اعضای یک خانواده تابع از تابع کوواریانس واقعی که به اندازه یک مقدار ثابت با هم تفاوت دارند، غیر قابل تمایز هستند. لذا هر یک از توابع ممکن را می توان بصورت:

$$\sigma(h) = C(h) + a \quad (12)$$

شامل یک قسمت مشترک $C(h)$ و یک قسمت ثابت a در نظر گرفت که در آن ثابت a موجب فرق گذاشتن یک تابع از بقیه می شود اما دارای اهمیت نیست. قسمت $C(h)$ را تابع کوواریانس تعمیم یافته (GCF) نامیده و در عمل هر تابعی را که فقط با یک مقدار ثابت از تابع دیگر فرق دارد می توان برای آن بکار برد. عنوان مثال، هر یک از توابع (۱۰) و (۱۱) را می توان GCF نامید.

برای کریگیدن عام نیز لزومی ندارد نگران پیدا کردن تابع کوواریانس کامل باشیم، بلکه در اختیار بودن GCF کافی

داده‌های روندزدوده بدست می‌آوریم. در اینصورت ماتریس $G\Sigma_R G^T$ کوواریانس نموهای تعمیم یافته، یعنی GR ، برابر خواهد بود. همچنین ماتریس کوواریانس نموهای تعمیم یافته داده‌های اصلی یعنی GZ نیز $G\Sigma G^T$ است و با استفاده از رابطه (۶۱) داریم:

$$G\Sigma_R G^T = G\Sigma G^T$$

بنابراین ماتریس‌های کوواریانس تعمیم یافته و در نتیجه توابع کوواریانس تعمیم یافته داده‌های اصلی و داده‌های روندزدوده یکسان هستند.

۴ نتیجه‌گیری

بدلیل وجود روند در داده‌ها، برآورد تابع کوواریانس برای کریگیدن عام، اریب است و در این حالت معمولاً آنرا با استفاده از تابع کوواریانس تجربی داده‌های روندزدوده بدست می‌آورند. برآورد حاصل از این روش که بستگی به نوع حذف روند دارد نیز اریب است. با توجه به اینکه داده‌های روندزدوده و داده‌های اصلی دارای توابع کوواریانس تعمیم یافته یکسانی هستند و کریگیدن حاصل از توابع کوواریانس معمولی و تعمیم یافته معادل یکدیگرند و از طرفی حذف روند از داده‌ها موجب حذف بیشتر تغییرات مربوط به روند در کوواریانس تجربی و آشکارتر شدن تابع کوواریانس تعمیم یافته می‌گردد، محاسبه کریگیدن عام بر اساس تابع کوواریانس تعمیم یافته داده‌های روندزدوده، توصیه می‌گردد.

و X یک ماتریس $(p+1) \times n$ با عناصر

$$x_{ij} = \phi_{j-1}(t_i) \quad i = 1, \dots, n; \quad j = 1, \dots, p+1$$

باشد. در اینصورت روند، به فرم:

$$\mu(t) = E(Z) = X\beta$$

خواهد بود و برآورد حداقل مربعات برای پارامتر β بصورت:

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Z$$

است و داده‌های بدون روند شده نیز بصورت:

$$R = Z - X\hat{\beta} = (I - X(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1})Z$$

محاسبه می‌گردد. بدیهی است میانگین داده‌های بدون روند شده برابر صفر است زیرا

$$\begin{aligned} E(R) &= (I - X(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1})E(Z) \\ &= (I - X(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1})X\beta = 0 \end{aligned}$$

و ماتریس کوواریانس آنها عبارتست از:

$$\begin{aligned} \Sigma_R &= E(RR^T) \\ &= \Sigma - X(X^T \Sigma^{-1} X)^{-1} X^T \end{aligned} \quad (۶)$$

همانطور که در (۶۱) ملاحظه می‌شود در حالت کلی Σ و Σ_R با یکدیگر برابر نیستند. اما تفاوت آنها مربوط به قسمت اصلی نیست. برای حذف قسمت غیر اصلی، نموهای تعمیم یافته ^{۱۱} را از ضرب ماتریسی مانند G ، که $0 = GX$ باشد، در

مراجع

- [1] Armstrong, M. (1984), Problems with Universal Kriging, *Mathematical Geology*, **16**, 101-108.
- [2] Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley and Sons, New York.
- [3] Cressie, N. and Gronoda, M. O. (1992), A Comparison of Variogram Estimation with Covariogram Estimation, In *The Art of Statistical Science*, K. V. Mardia, ed. Wiley, Chichester.
- [4] Christensen, R. (1990), *Linear Models for Multivariate, Time Series and Spatial Data*, Springer Verlag, New York.
- [5] Delfiner, P. (1976), Linear Estimation of Nonstationary Spatial Phnomena, In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Hjbregts, eds. Riedel, Dordrecht, 46-68.
- [6] Krige, D. G. (1951), A Statistical Approach to Some Basic Mine Valuation Problems on The Witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, **52**, 119-139.
- [7] Matheron, G. (1963), Principles of Geostatistics, *Economic Geology*, **58**, 1246-1266.
- [8] Matheron, G. (1969), Le Krigeag, Universal, *Cahiers du Center de Morphologie Mathematique*, No. 1 Fountainbleau, France.
- [9] Matheron, G. (1971), The Theory of Regionalized Variables and its Applications, *Cahiers du Center de Morphologie Mathematique* No. 5, Fountainbleau, France.
- [10] Matheron, G. (1973), The Intrinsic Random Functions and their Applications. *Adv. appl. prob.* **5**, 439-468.