

تابع کواریانس تعمیم یافته برای کریگیدن عام

محسن محمد زاده^۱

چکیده

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

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

۱ مقدمه

داده‌های فضایی^۲ 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 R^d ، $d \geq 1$ است و به ازاء هر t ، $Z(t)$ یک متغیر تصادفی با میانگین و واریانس

$$Var(Z(t)) = E[(Z(t) - \mu(t))^2] \quad (2)$$

است. همچنین برای دو موقعیت t_1 و t_2 در D تابع کواریانس $Z(t_1)$ و $Z(t_2)$ که بصورت:

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

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

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

ارائه خواهد شد، برآورد می‌شود. چون معمولاً تابع کوواریانس نامعلوم است، برآورد 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

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

برازش می‌شود. در عبارت (۵) مجموع روی تمام اعضای مجزای مجموعه $Nh = \{(i, j) : t_i - t_j = h\}$ گرفته می‌شود و h نشان دهنده فاصله دو موقعیت، $\bar{Z}$ یک برآوردگر میانگین $Z(\cdot)$ 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. بعلاوه کرسی و گروندونا (۱۹۹۲) برای این حالت تحت شرایط خاص جمع پذیری تابع کوواریانس نشان دادند برآورد تابع کوواریانس دارای اربیی از مرتبه $O(1/n)$ است. لذا اغلب برای برآورد تابع کوواریانس در کریگیدن عام، ابتدا روند را از داده‌ها حذف نموده، سپس همان رویه فوق بر داده‌های روند زده 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 اربیی در برآورد 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 برآورد تابع کوواریانس تعمیم یافته بکاربرد و قابل استفاده برای برآورد تابع

تعریف 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^۳ و یا روند^۴ و $\sigma(t)$ را قسمت تصادفی مدل (۴) می‌نامند. چنانچه میانگین و تابع کوواریانس میدان تصادفی بصورت:

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

باشند، یعنی میانگین مقداری ثابت و کوواریانس تابعی از فاصله دو موقعیت t_1 و t_2 باشند، آنرا میدان تصادفی مانای مرتبه دوم^۵ می‌نامند.

فرض کنید مشاهدات $Z = (Z(t_1), \dots, Z(t_n))^T$ از میدان تصادفی $Z(\cdot)$ در n موقعیت فضایی $t = (t_1, \dots, t_n)$ در اختیار باشد. یکی از متداولترین روشهای برآورد فضایی برای مقدار میدان تصادفی در یک موقعیت مشخص t_0 ، بهترین برآوردگر خطی نااریب (BLUP) بر اساس 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 مرتبه 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 ، مقدار $Z(t_0)$ که جزئیات آنها در بخش ۲

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_0 عبارتست از انتخاب یک تخمین گر خطی بصورت:

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

که بطور یکنواخت ناریب باشد، یعنی

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

باشد و میانگین مجذور خطاهای

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

را نسبت به λ کمینه نماید. شرط بطور یکنواخت ناریب (۹) معادل آنستکه $\lambda^T X = x_0^T$ باشد که در آن $x_0 = (\phi_0(t_0), \dots, \phi_n(t_0))^T$ است.

کریگیدن در نقطه t_0 بصورت (۸) 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 $\sigma_0 = (\sigma(t_0 - t_1), \dots, \sigma(t_0 - t_n))$ است، محاسبه می گردد. بعلاوه مقدار کمینه میانگین مجذور خطای برآورد برابر

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

می باشد. بدیهی است اگر $p = 0$ و $\phi_0(t) = 1$ اختیار شود، کریگیدن عام به کریگیدن عادی تبدیل می گردد.

کوواریانس معمولی نیستند. از طرفی ماترون (۱۹۷۱) نشان داده است برآوردهای حاصل از تمام توابع کوواریانس تعمیم یافته یکسان هستند. همچنین کریستنسن (۱۹۹۰) نشان داد برآوردهای حاصل از کریگیدن با استفاده از تابع کوواریانس معمولی و GCF معادل یکدیگرند. بنابراین کریگیدن عام را می توان بوسیله برآورد GCF که براساس داده های روندزوده بدست آورده می شود، انجام داد.

۲ کریگیدن

ماترون (۱۹۷۱) کریگیدن عادی را برای برآورد فضایی میدان تصادفی در یک نقطه مشخص t_0 براساس بردار مشاهدات Z و با فرض آنکه میانگین یا روند در مدل (۴) مقدار ثابت $\mu(t) = \mu$ 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 $\sum \lambda_i = 1$ باشد معرفی کرد، که شرط $\sum \lambda_i = 1$ ناریب بودن برآوردگر را تضمین می نماید و ضرایب برآوردگر نیز به گونه ای انتخاب می شوند که میانگین مجذور خطاها کمینه گردد.

در روش کریگیدن عام که توسط ماترون (۱۹۶۹) ارایه گردیده است فرض می شود روند در مدل (۴) یک ترکیب خطی از توابعی از قبل تعیین شده مانند $\phi_j(\cdot)$ 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 می شوند. دراین حالت مدل (۴) را می توان به فرم ماتریسی

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

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

۳ تابع کوواریانس تعمیم یافته

در روش کریگیدن عادی، معلوم بودن تابع کوواریانس تا یک مقدار ثابت کفایت می کند. بعنوان 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 ناریبی است، زیرا اضافه کردن یک مقدار ثابت به تابع کوواریانس معادل آنستکه به میدان تصادفی $Z(\cdot)$ یک میدان تصادفی با میانگین صفر و واریانس ثابت a اضافه نماییم. چون اضافه کردن یک مقدار ثابت به میدان تصادفی $Z(\cdot)$ در خطای برآورد بی تأثیر است، a هیچ تغییری در کریگیدن ایجاد نمی کند. پس برای کریگیدن عادی، تمام اعضای یک خانواده توابع از تابع کوواریانس واقعی که به اندازه یک مقدار ثابت باهم 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 توابع دیگر فرق دارد می توان برای آن بکار برد. بعنوان 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

^{۱۰} Nugget Effect

می باشد. در کریگیدن عام تعداد قیود ناریبی برابر تعداد ضرایب نامعلوم روند خطی، یعنی p در عبارت (۶) است. هرچقدر تعداد قیود ناریبی بیشتر باشد، تعداد توابع کوواریانسی که کاملاً یکسان عمل می کنند بیشتر می شود. بعنوان 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 (۱۵) که GCF را بعنوان 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 ناریبی، آزادی بیشتری برای انتخاب یک تابع کوواریانس فراهم می سازد. بطور کلی، با معرفی جملات بیشتر در روند یا قسمت قطعی مدل (۴)، عملاً 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^{۱۰} ساده تبدیل می گردد.

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

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

داده‌های روندزده بدست می‌آوریم. در اینصورت ماتریس کوواریانس نمونه‌های تعمیم یافته، یعنی GR ، برابر $G\Sigma_R G^T$ خواهد بود. همچنین ماتریس کوواریانس نمونه‌های 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 روندزده یکسان هستند.

۴ نتیجه گیری

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 روندزده بدست می‌آورند. برآورد حاصل از این روش که بستگی به نوع حذف روند دارد نیز اریب است. با توجه به اینکه داده‌های روندزده 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 روندزده، 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 می‌گردند. 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:

$$\begin{aligned} \Sigma_R &= E(RR^T) \\ &= \Sigma - X(X^T \Sigma^{-1} X)^{-1} X^T \end{aligned} \quad (16)$$

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، نمونه‌های تعمیم یافته^{۱۱} را از ضرب ماتریسی مانند G ، که $GX = 0$ باشد، در

- [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 Phenomena, In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Hijbregts, 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 Krigeage, 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.
-