ESTUDO DE SIMULAÇÃO DE MODELOS GEOESTATÍSTICOS MULTIVARIADOS

Bruno Henrique Fernandes Fonseca, Paulo Justiniano Ribeiro Jr.

1 REVISÃO

Bivariate gaussian common component model (BGCCM):

$$Y_1 = X_1\beta_{01} + \sigma_{01}R_{01} + \sigma_1 R_1$$
$$Y_2 = X_2\beta_{02} + \sigma_{02}R_{02} + \sigma_2 R_2$$

Parâmetros associados ao BGCCM, assumindo que as funções de correlações adotadas sejam todas de Matérn com κ fixo:

$$\theta = (\beta_{01}, \beta_{02}, \sigma_{01}, \sigma_2, \sigma_{02}, \sigma_2, \phi_0, \phi_1, \phi_2) \tag{1}$$

Função de verossimilhança associada a $(\theta; Y)$:

$$L(\theta; Y) \propto |\Sigma|^{-0.5} e^{-0.5(Y-X\beta)^t\Sigma^{-1}(Y-X\beta)};$$

sendo $\beta^t = (\beta_{01}, \beta_{02}$ e Σ depende de todos parâmetros exceto os de média.

Reparametrizações:

$$\sigma_{01} = \sigma,\ \sigma_1 = \sigma\nu_1,\ \sigma_{02} = \sigma\eta,\ \sigma_2 = \sigma\nu_2$$

logo:

$$\theta^* = (\beta, \sigma, \eta, \nu_1, \nu_2, \phi_0, \phi_1, \phi_2)$$

Função de verossimilhança associada a $(\theta^*; Y)$:

$$L(\theta^*; Y) \propto \sigma^{-n}|V|^{-0.5} e^{-0.5\sigma^{-2}(Y-X\beta)^t V^{-1}(Y-X\beta)}$$

sendo V uma matriz que depende de $\theta_c = (\eta, \nu_1, \nu_2, \phi_0, \phi_1, \phi_2)$.

Usando o logaritmo de $L(\theta^*; Y)$ para facilitar diferenciação:

$$l(\theta^*) = ln(L(\theta^*; Y)) \propto -nlog(\sigma) - 0.5log(|V|) - 0.5\sigma^{-2}(Y - X\beta)^t V^{-1}(Y - X\beta)$$

Estimador de θ^*:

$$\frac{\partial l(\theta^*)}{\partial \theta^*} = (0, 0, 0, 0, 0, 0, 0, 0)^t$$

Resolvendo as derivadas parciais com relação a β e σ e igualando as mesmas a zero tem-se:

$$\widehat{\beta} = (X^t V^{-1} X)^{-1}(X^t V^{-1} Y) \;\; \widehat{\sigma} = \sqrt{n^{-1}(Y - X\beta)^t V^{-1}(Y - X\beta)}$$

Agora fazendo $T_\beta = \frac{\partial^2 l(\theta^*)}{\partial \beta^2}$ e $T_\sigma = \frac{\partial^2 l(\theta^*)}{\partial \sigma^2}$ e utilizando as propriedades assintóticas dos estimadores de máxima verossimilhança, tem-se:

$$\widehat{\beta} \backsim N(\beta; I(\beta)^{-1}) \;\; \widehat{\sigma} \backsim N(\sigma; I(\sigma)^{-1})$$

sendo $I(\beta) = E(-T_\beta)$ e $I(\beta) = E(-T_\sigma)$, que são as informações de Fischer esperadas, porém iremos utilizar as observadas.

PRIMEIRA DÚVIDA

As derivadas segundas cruzadas entre β e σ possuem forma analítica, logo, eu devo considerar a distribuição conjunta entre os mesmos??? Ou posso calcular as variâncias marginalmente conforme exposto acima???

Com relação aos demais parâmetros de θ^* as derivadas parciais $l(\theta^*)$ não possuem forma analítica de fácil resolução, sendo assim, métodos numéricos de maximização serão utilizados. Além disso, usando $\widehat{\beta}$ e $\widehat{\sigma}$ tem-se o logaritmo da função de verossimilhança concentrada:

$$l(\theta_c) \propto -nlog(Q) - 0.5log(|V|)$$

sendo $Q = (Y - X\widehat{\beta})^t V^{-1}(Y - X\widehat{\beta})$.

Para calcular as estimativas de θ_c foi utilizado o algotirítimo de Nelder-Mead. Que resulta o $\widehat{\theta}_c$ e a matriz hessiana estimada que é $T_{\theta_c} = \frac{\widehat{\partial^2 l(\theta_c)}}{\partial \theta_c^2}$.

Sendo assim, utilizando as propriedades assintóticas dos estimadores de máxima verossimilhança tem-se:

$$\widehat{\theta_c} \backsim N(\theta_c; T(\theta_c)^{-1}),$$

Observação: na distribuição acima já foi plugada a estimativa da matriz de informação de Fischer observada.

Sendo assim, podemos calcular as estimativas das médias e variâncias de $\widehat{\beta}$ e $\widehat{\sigma}$.

No entando, devemos recuperar também a distribuição dos parâmetros originais θ, logo, iremos utilizar o método delta para tal:

$$\widehat{\theta} \backsim N(\theta; \Sigma_\theta),$$

sendo $\Sigma_\theta = \Delta^t \Sigma_{\theta^*} \Delta$, onde:

$$\Sigma_{\theta^*} = \begin{bmatrix} \Sigma_{\beta,\sigma} & O \\ O^t & \Sigma_{\theta_c} \end{bmatrix}$$

sendo $\Sigma_{\beta,\sigma}$ a matriz 3×3 de covariâncias para β e σ, Σ_{θ_c} a matriz 6×6 de covariâncias de θ_c e O uma matriz 3×6 de zeros.

A i-ésima coluna de Δ é igual a $\frac{\partial \theta_i}{\partial \theta^*}$, para i = 1, 2, ..., 9.

SEGUNDA DÚVIDA

Realmente devo utilizar Σ_{θ^*} para calcular Σ_θ??? Se sim, é correto plugar a matriz O na mesma??? Pois assim estaremos supondo independência entre (β, σ) e θ_c. Ou devo utilizar apenas a matriz de Σ_c para calcular as variâncias de σ_{02}, σ_1 e σ_2. Por exemplo:

$$var(\sigma_{02}) = var(g(\eta)) = \frac{\partial \sigma_{02}^t}{\partial \theta_c} \Sigma_c \frac{\partial \sigma_{02}}{\partial \theta_c} = (\sigma, 0, 0, 0, 0, 0) \Sigma_c (\sigma, 0, 0, 0, 0, 0)^t$$

porém calculando a variância de $\sigma_{02} = \sigma\eta$ dessa forma, estaremos considerando que σ_{02} é função apenas de η, logo estaremos considerando que σ_{02} é independente de σ.

Para avaliar a precisão das estimativas, iremos utilizar intervalos de confiança marginais, por exemplo:

$$IC(\beta_{01}, 0.95) = \widehat{\beta_{01}} \pm 1.96\sqrt{var(\beta_{01})}$$

Porém no contexto de simulação, teremos 1000 intervalos de confiança para cada parâmetro, os quais devem possuir em sua maioria os parâmetros utilizados nas simulações.

Quando forem simulados dados com BGCCM serão armazenados 20 pontos georeferenciados e os valores das respotas, ou seja, teremos esses dados simulados, porém na modelagem eles não serão considerados. Com os parâmetros estimados fizemos a krigagem espacial nos 20 pontos não utilizados. Sendo assim, teremos 20 valores preditos para cada resposta e seus respectivos valores simulados. Sendo assim, novamente podemos pensar no contexto dos intervalos de confiança, onde agora os intervalos preditivos devem conter o valor simulado.

TERCEIRA DÚVIDA

Como temos 20 intervalos de confiança preditivos para cada resposta em cada uma das mil simulações. Como faço para avaliar de forma mais sucinta???

2 PRIMEIROS RESULTADOS

2.1 CONFIGURAÇÃO

A primeira configuração utilizada foi: $n_1 = 50$, $n_2 = 50$, grid irregular em um quadrado de lado igual a 1, localizações co-locadas, $\mu_1 = 150$, $\mu_2 = 60$, $\sigma_{01} = 8$, $\sigma_1 = 4$, $\sigma_{02} = 5$, $\sigma_2 = 2$, $\phi_0 = \phi_1 = \phi_2 = 0.2$, $\kappa_0 = 1.5$, $\kappa_1 = 0.5$ e $\kappa_2 = 0.5$. Observação: só utilizarei simulações que gerem respostas com 0.7 ou mais de correlação marginal.

2.2 RESULTADOS

Para ilustrar as análises dos resultados do estudo de simulação, inicialmente, mostraremos os resultados para uma única amostra da configuração em questão. Coordenadas:

Dados:

Y_1 = (143,500 168,090 142,131 152,431 146,973 158,714 149,247 142,429 141,922 152,143 146,605 146,846 147,203 137,711 139,678 139,895 142,620 139,716 151,113 148,045 141,466 149,989 142,895 159,900 133,613 154,845 136,554 141,511 139,537 145,466 155,980 154,990 142,962 162,284 130,919 141,616 138,624 141,200 160,989 143,551 145,748 151,544 136,277 142,816 138,207 139,297 139,893 134,216 143,011 153,931)

Y_2 = (56,410 68,190 54,939 58,709 52,243 66,027 57,754 55,810 51,658 58,344 57,742 59,506 54,514 52,581 57,218 62,251 54,556 57,910 63,232 60,122 60,321 61,835 54,751 66,192 51,573 62,996 58,526 52,995 58,013 53,453 64,184 65,063 52,556 68,034 48,857 55,314 55,178 49,879 66,430 56,599 58,862 58,390 57,970 52,294 56,203 59,403 57,931 51,879 57,436 62,569)

A correlação marginal entre Y_1 e Y_2 é de 0.793.

Resumo das estimativas:

θ	Estimativa	Erro Padrão	$IC_{0.95}$
β_{01}	$145,90$	$4,27$	$(137,53;154,27)$
β_{02}	$57,49$	$2,56$	$(52,48;62,51)$
σ_{01}	$8,44$	$0,60$	$(7,27;9,61)$
σ_1	$3,06$	$1,18$	$(0,74;5,39)$
σ_{02}	$5,06$	$0,81$	$(3,47;6,65)$
σ_2	$1,90$	$0,66$	$(0,62;3,18)$
ϕ_0	$0,18$	$0,05$	$(0,08;0,28)$
ϕ_1	$0,12$	$0,06$	$(0,005;0,23)$
ϕ_2	$0,11$	$0,05$	$(0,002;0,21)$

TABELA 1: EMV - BGCCM

Os resultados acima mostram que todas as estimativas intervalares marginais contemplam os parâmetros originais. O próximo passo é analisar as predições espaciais:

$Y_{1,arm}$	$Y_{1,pred}$	Erro Padrão	$IC_{0.95}$
$137,334$	$139,037$	$9,819$	$(118,088;156,580)$
$145,698$	$140,881$	$9,413$	$(127,249;164,148)$
$145,214$	$147,148$	$9,055$	$(127,465;162,962)$
$146,787$	$146,218$	$9,580$	$(128,009;165,564)$
$149,809$	$149,667$	$9,534$	$(131,121;168,496)$
$151,230$	$153,217$	$9,107$	$(133,380;169,080)$
$141,644$	$142,573$	$9,591$	$(122,846;160,442)$
$142,061$	$140,826$	$9,798$	$(122,857;161,266)$
$152,856$	$152,972$	$9,111$	$(134,999;170,713)$
$148,527$	$147,669$	$9,496$	$(129,914;167,140)$
$134,938$	$136,477$	$9,372$	$(116,570;153,306)$
$128,792$	$133,096$	$9,329$	$(110,507;147,078)$
$129,679$	$133,011$	$9,316$	$(111,420;147,937)$
$143,611$	$142,115$	$9,164$	$(125,650;161,571)$
$149,175$	$150,129$	$9,598$	$(130,363;167,987)$
$143,475$	$141,452$	$9,295$	$(125,258;161,693)$
$146,024$	$150,155$	$0,322$	$(125,793;166,255)$
$156,412$	$155,405$	$9,232$	$(138,317;174,507)$
$146,897$	$146,556$	$9,170$	$(128,924;164,870)$
$137,284$	$138,734$	$9,921$	$(117,839;156,729)$

TABELA 2: Krigagem nos pontos retirados de Y_1

$Y_{1,arm}$	$Y_{1,pred}$	Erro Padrão	$IC_{0.95}$
52.409	52.358	2.601	(47.311; 57.507)
53.222	54.530	1.981	(49.339; 57.104)
53.753	54.333	1.177	(51.446; 56.060)
59.337	57.748	2.253	(54.921; 63.752)
57.280	58.035	2.182	(53.003; 61.556)
61.428	60.145	1.326	(58.829; 64.027)
54.934	54.496	2.272	(50.481; 59.387)
53.350	53.769	2.572	(48.309; 58.391)
61.446	60.297	1.336	(58.828; 64.065)
60.973	58.225	2.117	(56.823; 65.122)
51.000	52.080	1.901	(47.274; 54.726)
49.519	49.988	1.827	(45.938; 53.100)
49.686	49.939	1.801	(46.156; 53.216)
57.098	57.080	1.475	(54.207; 59.989)
60.454	60.111	2.285	(55.976; 64.933)
53.775	53.852	1.760	(50.325; 57.224)
56.711	59.653	3.224	(50.392; 63.030)
63.798	63.686	1.625	(60.613; 66.983)
63.343	60.966	1.492	(60.419; 66.267)
52.452	53.030	2.737	(47.087; 57.816)

TABELA 3: Krigagem nos pontos retirados de Y_2

Para as duas variáveis os valores preditos ficaram próximos aos observados e não utilizados na estimação, além disso todos intervalos de confiança contemplam os valores observados. No entanto, além de analisar os resultados para uma única amostra das variáveis respostas, nós conduzimos um estudo de simulação com 100 amostras do BGCCM com as configurações paramétricas em questão. A tabela abaixo, ilustra os resultados:

β_{01}	β_{02}	σ_{01}	σ_1	σ_{02}	σ_2	ϕ_0	ϕ_1	ϕ_2
925	922	426	919	834	646	892	769	769

TABELA 4: Contagem dos intervalos de confiança para θ que contemplam os parâmetros de simulação