

Block Properties for Biwm

Leonardo Uchoa

Todas as contas a seguir tem o intuito de melhorar problemas de instabilidade numérica, reduzir o custo computacional de inversas e simplesmente fazer este belíssimo código funcionar.

Contas na Log-Verossimilhança

Log do Determinante

Se

$$\Sigma_{\theta}(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & C_{12}(\mathbf{h}) \\ C_{21}(\mathbf{h}) & C_{22}(\mathbf{h}) \end{pmatrix} \quad (1)$$

Então

$$\det(\Sigma) = |\Sigma| = \left| C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right| * \left| C_{22}(\mathbf{h}) \right| \quad (2)$$

Portanto, com o logaritmo fica

$$\log(\det(\Sigma)) = \log \left| C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right| + \log \left| C_{22}(\mathbf{h}) \right| \quad (3)$$

Extra

Para o nosso caso também é válido que

$$\det(\Sigma) = \left| C_{11}(\mathbf{h})C_{22}(\mathbf{h}) - C_{12}(\mathbf{h})C_{21}(\mathbf{h}) \right|$$

pois as matrizes $C_{12}(\mathbf{h})$ e $C_{22}(\mathbf{h})$ comutam. Além disso como $C_{12}(\mathbf{h}) = C_{21}(\mathbf{h})$

$$\det(\Sigma) = \left| C_{11}(\mathbf{h})C_{22}(\mathbf{h}) - C_{12}^2(\mathbf{h}) \right|$$

Inversa

Aqui existem duas formas. Eu preferi usar a que irei apresentação a seguir pois ela encaixa bem com o resultado (3) do log do determinante¹.

Se

¹Apesar de eu achar que não tem como o alimentar o optim com as inversas obtidas, para ganhar tempo

$$\Sigma_{\theta}^{-1}(\mathbf{h}) = \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \quad (4)$$

Então

$$\begin{aligned} C_{11}^*(\mathbf{h}) &= \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} \\ C_{12}^*(\mathbf{h}) &= - \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) &= -C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} \\ C_{22}^*(\mathbf{h}) &= C_{22}^{-1}(\mathbf{h}) + C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \left[C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \right]^{-1} C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}) \end{aligned} \quad (5)$$

Já que² $C_{12}(\mathbf{h}) = C_{21}(\mathbf{h})$, se definirmos

$$\begin{aligned} V_1 &:= C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \\ V_2 &:= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h}) \end{aligned} \quad (6)$$

podemos obter uma forma simplificada para o sistema (5), já que como $C_{11}, C_{12}, C_{21}, C_{22}$ são simétricas

$$\begin{aligned} C_{11}^*(\mathbf{h}) &= V_2^{-1} \\ C_{12}^*(\mathbf{h}) &= -V_2^{-1}V_1^T \\ C_{21}^*(\mathbf{h}) &= -V_1V_2^{-1} \\ C_{22}^*(\mathbf{h}) &= C_{22}^{-1}(\mathbf{h}) - V_1V_2^{-1}V_1^T \end{aligned} \quad (7)$$

pois $V_1 = C_{22}^{-1}(\mathbf{h})C_{21}(\mathbf{h})$ fornece $V_1^T = C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})$.

Note ainda que $C_{12}^*(\mathbf{h}) = C_{12}^{*T}(\mathbf{h})$.

Prova:

Primeiramente

$$\begin{aligned} V_2^T &= C_{11}(\mathbf{h}) - (C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{12}(\mathbf{h}))^T \\ &= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})(C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h}))^T \\ &= C_{11}(\mathbf{h}) - C_{12}(\mathbf{h})C_{22}^{-1}(\mathbf{h})C_{12}(\mathbf{h}) \\ &= V_2 \end{aligned} \quad (8)$$

Agora

$$\begin{aligned} C_{12}^{*T}(\mathbf{h}) &= -(V_2^{-1}V_1^T)^T \\ &= V_1^T(V_2^{-1})^T \\ &= V_1^T(V_2^{-1}) \\ &= C_{21}^*(\mathbf{h}) \end{aligned} \quad (9)$$

²No artigo Σ é definida assim

Avaliação de Σ^{-1} na log-verossimilhança

Neste caso precisamos avaliar $y^T \Sigma^{-1} y$. A estrutura de blocos também pode nos ajudar a deixar o algoritmo um pouco mais estável e rápido.

Novamente a inversa de Σ é denotada por

$$\Sigma_{\theta}^{-1}(\mathbf{h}) = \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \quad (10)$$

Agora se $y^T = (y_1, y_2)^T$, em que y_1 e y_2 são, respectivamente, os vetores resposta associados às primeira e segunda variáveis multivariadas, temos que

$$\begin{aligned} y^T \Sigma_{\theta}^{-1}(\mathbf{h}) y &= y^T \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} y \\ &= (y_1, y_2)^T \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= (y_1, y_2)^T \begin{pmatrix} C_{11}^*(\mathbf{h})y_1 + C_{12}^*(\mathbf{h})y_2 \\ C_{21}^*(\mathbf{h})y_1 + C_{22}^*(\mathbf{h})y_2 \end{pmatrix} \\ &= y_1^T C_{11}^*(\mathbf{h})y_1 + y_1^T C_{12}^*(\mathbf{h})y_2 + y_2^T C_{21}^*(\mathbf{h})y_1 + y_2^T C_{22}^*(\mathbf{h})y_2 \\ &= y_1^T C_{11}^*(\mathbf{h})y_1 + 2y_1^T C_{12}^*(\mathbf{h})y_2 + y_2^T C_{22}^*(\mathbf{h})y_2 \end{aligned} \quad (11)$$

pois como $C_{12}^{*T}(\mathbf{h}) = C_{21}^*(\mathbf{h})$ obtemos que

$$y_1^T C_{12}^*(\mathbf{h})y_2 = \text{tr}(y_1^T C_{12}^*(\mathbf{h})y_2) = \text{tr}(y_1^T C_{12}^*(\mathbf{h})y_2)^T = \text{tr}((C_{12}^*(\mathbf{h})y_2)^T y_1) = \text{tr}(y_2^T C_{12}^{*T}(\mathbf{h})y_1) = y_2^T C_{21}^*(\mathbf{h})y_1$$

Também é possível escrever o resultado anterior em termos de (11), o que fornece

$$y_1^T V_2^{-1} y_1 - 2y_1^T V_2^{-1} V_1^T y_2 + y_2^T C_{22}^{-1} y_2 - y_2^T V_1 V_2^{-1} V_1^T y_2$$

Contas no Gradiente

Aqui a idéia é mitigar o custo computacional de $\text{tr}\left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \theta_k}\right)$ já que vão aparecer muitas matrizes nulas bem grandes em $\partial \Sigma / \partial \theta_k$

Relembrar é viver: Produto de Matrizes

Se

$$\Sigma_{\theta}(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & C_{12}(\mathbf{h}) \\ C_{21}(\mathbf{h}) & C_{22}(\mathbf{h}) \end{pmatrix} \quad (12)$$

e

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad (13)$$

Então desde as matrizes M e $\Sigma_{\theta}(\mathbf{h})$ sejam conformes

$$\Sigma_{\theta}(\mathbf{h})M = \begin{pmatrix} C_{11}(\mathbf{h})A + C_{12}(\mathbf{h})C & C_{11}(\mathbf{h})B + C_{12}(\mathbf{h})D \\ C_{21}(\mathbf{h})A + C_{22}(\mathbf{h})C & C_{21}(\mathbf{h})B + C_{22}(\mathbf{h})D \end{pmatrix} \quad (14)$$

Traços: $tr \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \theta_k} \right)$

Caso em que $\theta_k = \sigma_1^2$

$$\begin{aligned} \Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \sigma_1^2} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} M(\mathbf{h}|\nu_1, a) & \mathbf{O} \\ \frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & \mathbf{O} \end{pmatrix} \\ &= \begin{pmatrix} C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \\ C_{21}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{11}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \end{pmatrix} \end{aligned} \quad (15)$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned} tr \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \sigma_1^2} \right) &= tr \left(C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) \right) + 2tr \left(C_{12}^*(\mathbf{h})\frac{\rho\sigma_2}{2\sigma_1} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \right) \\ &= tr \left(C_{11}^*(\mathbf{h})M(\mathbf{h}|\nu_1, a) \right) + \frac{\rho\sigma_2}{\sigma_1} tr \left(C_{12}^*(\mathbf{h})M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \right) \end{aligned} \quad (16)$$

Caso em que $\theta_k = \sigma_2^2$

$$\begin{aligned} \Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \sigma_2^2} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \mathbf{O} & M(\mathbf{h}|\nu_2, a) \\ \frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & M(\mathbf{h}|\nu_2, a) \end{pmatrix} \\ &= \begin{pmatrix} C_{11}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{11}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) + C_{12}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a) \\ C_{22}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{21}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) + C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a) \end{pmatrix} \end{aligned} \quad (17)$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned} tr \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \sigma_2^2} \right) &= tr \left(C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a) \right) + 2tr \left(C_{12}^*(\mathbf{h})\frac{\rho\sigma_1}{2\sigma_2} M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \right) \\ &= tr \left(C_{22}^*(\mathbf{h})M(\mathbf{h}|\nu_2, a) \right) + \frac{\rho\sigma_1}{\sigma_2} tr \left(C_{12}^*(\mathbf{h})M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \right) \end{aligned} \quad (18)$$

Caso em que $\theta_k = \rho$

$$\begin{aligned} \Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \rho} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \sigma_1\sigma_2 M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & \mathbf{O} \end{pmatrix} \\ &= \begin{pmatrix} C_{12}^*(\mathbf{h})\sigma_1\sigma_2 M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{11}^*(\mathbf{h})\sigma_1\sigma_2 M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \\ C_{22}^*(\mathbf{h})\sigma_1\sigma_2 M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{21}^*(\mathbf{h})\sigma_1\sigma_2 M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \end{pmatrix} \end{aligned} \quad (19)$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$tr \left(\Sigma_{\theta}^{-1}(\mathbf{h}) \frac{\partial \Sigma}{\partial \rho} \right) = 2\sigma_1\sigma_2 tr \left(C_{22}^*(\mathbf{h})M(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \right) \quad (20)$$

Caso em que $\theta_k = a$

$$\begin{aligned}\Sigma_{\theta}^{-1}(\mathbf{h})\frac{\partial\Sigma}{\partial a} &= \begin{pmatrix} C_{11}^*(\mathbf{h}) & C_{12}^*(\mathbf{h}) \\ C_{21}^*(\mathbf{h}) & C_{22}^*(\mathbf{h}) \end{pmatrix} \begin{pmatrix} \sigma_1^2 M'(\mathbf{h}|\nu_1, a) & \rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) \\ \rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & \sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix} \\ &= \begin{pmatrix} C_{11}^*(\mathbf{h})\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + C_{12}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{11}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) + C_{12}^*(\mathbf{h})\sigma_2^2 M'(\mathbf{h}|\nu_2, a) \\ C_{21}^*(\mathbf{h})\sigma_1^2 M'(\mathbf{h}|\nu_1, a) + C_{22}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) & C_{21}^*(\mathbf{h})\rho\sigma_1\sigma_2 M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a) + C_{22}^*(\mathbf{h})\sigma_2^2 M'(\mathbf{h}|\nu_2, a) \end{pmatrix} \end{aligned} \quad (21)$$

Portanto como $C_{12}^*(\mathbf{h}) = C_{21}^{*T}(\mathbf{h})$

$$\begin{aligned}tr\left(\Sigma_{\theta}^{-1}(\mathbf{h})\frac{\partial\Sigma}{\partial a}\right) &= \sigma_1^2 tr\left(C_{11}^*(\mathbf{h})M'(\mathbf{h}|\nu_1, a)\right) + \\ &\quad \sigma_2^2 tr\left(C_{22}^*(\mathbf{h})M'(\mathbf{h}|\nu_2, a)\right) + \\ &\quad 2\rho\sigma_1\sigma_2 tr\left(C_{12}^*(\mathbf{h})M'(\mathbf{h}|\frac{\nu_1+\nu_2}{2}, a)\right) \end{aligned} \quad (22)$$

Por outro lado, temos que

$$M'(\mathbf{h}|\nu, a) = \frac{\partial M(\mathbf{h}|\nu, a)}{\partial a} = \frac{(\frac{2}{a})^{1-\nu} d^\nu}{\Gamma(\nu)} \left[2\nu K_\nu(ad) - adK_{\nu+1}(ad) \right] \quad (23)$$

Se definirmos

$$\begin{aligned}\phi_i &:= \frac{(\frac{2}{a})^{1-\nu_i} d^{\nu_i}}{\Gamma(\nu_i)} \\ \Phi_i &:= \left[2\nu_i K_{\nu_i}(ad) - adK_{\nu_i+1}(ad) \right] \end{aligned} \quad (24)$$

e desenvolvermos um pouco mais as contas, temos que

$$\begin{aligned}tr\left(\Sigma_{\theta}^{-1}(\mathbf{h})\frac{\partial\Sigma}{\partial a}\right) &= \sigma_1^2 \phi_1 tr\left(C_{11}^*(\mathbf{h})\Phi_1\right) + \\ &\quad \sigma_2^2 \phi_2 tr\left(C_{22}^*(\mathbf{h})\Phi_2\right) + \\ &\quad 2\rho\sigma_1\sigma_2 \phi_3 tr\left(C_{12}^*(\mathbf{h})\Phi_3\right) \end{aligned} \quad (25)$$