

Reparametrização em modelos de
regressão não linear: um enfoque
Bayesiano

Jorge Alberto Achcar

nº 43

São Carlos
1989

REPARAMETRIZAÇÃO EM MODELOS DE REGRESSÃO NÃO-LINEAR: UM ENFOQUE BAYESIANO

JORGE ALBERTO ACHCAR

SUMÁRIO

O problema de boa parametrização é essencial para se conseguir bons resultados em modelos não-lineares. Por exemplo, um procedimento numérico iterativo para determinar estimadores de mínimos quadrados pode não convergir numa parametrização, mas pode convergir rapidamente numa parametrização adequada. Recentemente, vários artigos são apresentados na literatura relacionados à escolha de uma boa parametrização (ver por exemplo, Bates e Watts, JRSS, B, 42(1), 1-25, 1980; Cox e Reid, JRSS, B, 49(1), 1-39, 1987; Hougaard, JRSS, B, 44(2), 244-252, 1982). Neste artigo, apresentamos um enfoque Bayesiano para se achar uma boa parametrização onde a verossimilhança seja aproximadamente normal inclusive para amostras pequenas. Este procedimento envolve a solução de algumas equações diferenciais para se encontrar uma parametrização cuja priori não-informativa de Jeffreys seja localmente uniforme. Como ilustração, apresentamos alguns exemplos numéricos.

1. INTRODUÇÃO

Uma boa parametrização em modelos estatísticos é um problema de grande interesse prático. Sob o ponto de vista Bayesiano, observamos que uma densidade a posteriori é bem comportada sob uma parametrização, mas tem péssima forma sob outras parametrizações. Isto pode implicar em grandes erros quando utilizamos um método numérico ou um método aproximado na determinação de momentos e marginais a posteriori de interesse (ver por exemplo, Achcar e Smith, 1988).

Em modelos de regressão não-linear, observamos que uma boa parametrização é dada quando a verossimilhança se aproxima de uma densidade normal independentemente do tamanho da amostra considerada. Em termos práticos, isto pode levar à uma grande simplificação numérica dos métodos iterativos considerados para estimar os parâmetros e grande precisão. Por exemplo, considerando um modelo não-linear com dois parâmetros, podemos encontrar uma reparametrização onde obtemos contornos próximos de elipses para os parâmetros envolvidos.

A escolha de uma boa parametrização é amplamente discutida na literatura (ver por exemplo, Cox e Reid, 1987; Sprott, 1973, 1980; Diccio, 1984; Bates e Watts, 1980; Beale, 1960; Box, 1971; Holland, 1973). Em relação à escolha de uma boa parametrização em modelos não-lineares, podemos mencionar os artigos de Hougaard, 1982, 1984.

Neste artigo, discutimos o problema de boa parametrização em modelos não-lineares sob o enfoque Bayesiano. Assim, quando é possível resolver algumas equações diferenciais envolvidas com uma reparametrização com uma priori de Jeffreys localmente uniforme, teremos uma boa parametrização que coincide com as soluções sugeridas por Hougaard, 1982, 1984.

2. REPARAMETRIZAÇÃO NO CASO UNIPARAMÉTRICO

Supor o modelo de regressão não-linear envolvendo apenas um parâmetro θ ,

$$Y_i = \eta_i(\theta) + \epsilon_i \quad (1)$$

onde $i=1,2,\dots,n$, ϵ_i são independentes identicamente distribuídos com uma densidade normal $N(0, \sigma^2)$ e supor σ^2 conhecido.

O logaritmo da verossimilhança para θ , baseado numa amostra de tamanho n é dado por:

$$L(\theta) \propto -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \eta_i(\theta))^2 \quad (2)$$

As duas primeiras derivadas de $L(\theta)$ são dadas por,

$$\sigma^2 \frac{dL}{d\theta} = \sum_{i=1}^n \dot{\eta}_i(\theta) (y_i - \eta_i(\theta))$$

$$\sigma^2 \frac{d^2L}{d\theta^2} = -\sum_{i=1}^n \ddot{\eta}_i^2(\theta) + \sum_{i=1}^n \ddot{\eta}_i(\theta) (y_i - \eta_i(\theta)),$$

onde $\dot{\eta}_i$ e $\ddot{\eta}_i$ são as duas primeiras derivadas de $\eta_i(\theta)$.

Como $E(Y_i) = \eta_i(\theta)$, temos

$$E\left\{-\frac{d^2L}{d\theta^2}\right\} = \frac{1}{\sigma^2} \sum_{i=1}^n \ddot{\eta}_i^2(\theta) \quad (3)$$

A priori não-informativa de Jeffreys para θ é dada por,

$$\pi(\theta) \propto \{\underline{\eta}'(\theta)\underline{\eta}(\theta)\}^{1/2} \quad (4)$$

onde $\underline{\eta}'(\theta) = (\eta_1(\theta), \dots, \eta_n(\theta))$ e $\underline{\eta}(\theta) = (\dot{\eta}_1(\theta), \dots, \dot{\eta}_n(\theta))$.

Uma boa parametrização pode ser obtida quando podemos encontrar uma transformação um-a-um $\beta = g(\theta)$ que induz a uma priori de Jeffreys localmente uniforme $\pi(\beta) \propto \text{constante}$ (ver Achcar e Smith, 1988). Portanto,

$$\pi(\theta) \propto \text{constante} \left| \frac{dg}{d\theta} \right|, \quad (5)$$

ou $g'(\theta) \propto \{\underline{\eta}(\theta)\underline{\eta}(\theta)\}^{1/2}$, que é a boa parametrização sugerida por Hougaard, 1984 e por Bates e Watts, 1980.

3. REPARAMETRIZAÇÃO NO CASO MULTIPARAMÉTRICO

Supor o modelo não-linear (1) com θ substituído pelo vetor paramétrico $\underline{\theta}' = (\theta_1, \theta_2, \dots, \theta_k)$. Representando as primeiras derivadas parciais $\partial \eta_i(\underline{\theta}) / \partial \theta_j$ por $\dot{\eta}_{ij}(\underline{\theta})$, $i=1, 2, \dots, n$; $j=1, 2, \dots, k$, temos:

$$\frac{\partial L}{\partial \theta_j} = \frac{1}{\sigma^2} \{ \underline{y}' \dot{\eta}_{\cdot j}(\underline{\theta}) - \dot{\eta}'_j(\underline{\theta}) \underline{\eta}(\underline{\theta}) \} \quad (6)$$

onde $L(\underline{\theta})$ é o logaritmo da verossimilhança para $\underline{\theta}$, $\underline{y}' = (y_1, \dots, y_n)$, $\underline{\eta}(\underline{\theta}) = (\eta_1(\underline{\theta}), \dots, \eta_n(\underline{\theta}))$ e $\dot{\eta}'_j(\underline{\theta}) = (\dot{\eta}_{1j}(\underline{\theta}), \dots, \dot{\eta}_{nj}(\underline{\theta}))$ para $j=1, 2, \dots, k$.

Da mesma forma, calculamos as segundas derivadas parciais de

$L(\underline{\theta})$,

$$\frac{\partial^2 L}{\partial \theta_\ell \partial \theta_j} = \frac{1}{\sigma^2} \{ \underline{y}' \ddot{\eta}_{\ell j}(\underline{\theta}) - \dot{\eta}'_j(\underline{\theta}) \dot{\eta}'_\ell(\underline{\theta}) - \eta'(\underline{\theta}) \ddot{\eta}_{\ell j}(\underline{\theta}) \} \quad (7)$$

onde $\ell, j=1, 2, \dots, k$, e $\ddot{\eta}_{\ell j} = (\ddot{\eta}_{1\ell j}, \dots, \ddot{\eta}_{n\ell j})$.

Como $E(\underline{y}) = \eta(\underline{\theta})$, temos:

$$E \left\{ - \frac{\partial^2 L}{\partial \theta_\ell \partial \theta_j} \right\} = \frac{1}{\sigma^2} \dot{\eta}'_j(\underline{\theta}) \dot{\eta}'_\ell(\underline{\theta}) \quad (8)$$

Portanto, a matriz de informação de Fisher para $\underline{\theta}$ é dada por:

$$I(\underline{\theta}) = \frac{1}{\sigma^2} \begin{pmatrix} \dot{\eta}'_1 \dot{\eta}'_1 & \dot{\eta}'_1 \dot{\eta}'_2 & \dots & \dot{\eta}'_1 \dot{\eta}'_k \\ & \dot{\eta}'_2 \dot{\eta}'_2 & \dots & \dot{\eta}'_2 \dot{\eta}'_k \\ & & \ddots & \vdots \\ & & & \dot{\eta}'_k \dot{\eta}'_k \end{pmatrix} \quad (9)$$

simétrica

A priori não-informativa de Jeffreys para $\underline{\theta}$ (ver Box e Tiao, 1973) é dada por:

$$\pi(\underline{\theta}) \propto \{ \det I(\underline{\theta}) \}^{1/2} \quad (10)$$

Seja $\underline{\beta}' = (\beta_1, \dots, \beta_k) = (g_1(\underline{\theta}), \dots, g_k(\underline{\theta}))$ uma transformação de $\underline{\theta}$ tal que a priori implícita de Jeffreys seja $\pi(\underline{\beta}) \propto$ constante. Portanto, na parametrização $\underline{\theta}$, temos,

$$\pi(\underline{\theta}) \propto \text{constante } |J| \quad (11)$$

onde J é o jacobiano da transformação de $\underline{\beta}$, dado por,

$$J = \det \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \dots & \frac{\partial g_1}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial \theta_1} & \frac{\partial g_k}{\partial \theta_2} & \dots & \frac{\partial g_k}{\partial \theta_k} \end{pmatrix} \quad (12)$$

Portanto, uma boa parametrização $\underline{\beta}' = (g_1(\underline{\theta}), \dots, g_k(\underline{\theta}))$ é dada satisfazendo a equação diferencial $J = \text{constante } \{ \det I(\underline{\theta}) \}^{1/2}$, que implica numa priori de Jeffreys localmente uniforme para $\underline{\beta}$.

Como um caso especial, considere $\underline{\theta} = (\theta_1, \theta_2)$. A priori não-informativa de Jeffreys (10) para θ_1 e θ_2 é dada por:

$$\pi(\theta_1, \theta_2) \propto \{(\dot{\eta}'_1 \dot{\eta}_1)(\dot{\eta}'_2 \dot{\eta}_2) - (\dot{\eta}'_1 \dot{\eta}'_2)^2\}^{1/2} \quad (13)$$

onde $\dot{\eta}_j = \dot{\eta}_j(\theta)$ e $\theta' = (\theta_1, \theta_2)$.

Então, uma boa parametrização $(g_1(\theta), g_2(\theta))$ satisfaz a equação diferencial,

$$\begin{aligned} \frac{\partial g_1(\theta)}{\partial \theta_1} \frac{\partial g_2(\theta)}{\partial \theta_2} - \frac{\partial g_1(\theta)}{\partial \theta_2} \frac{\partial g_2(\theta)}{\partial \theta_1} &= \\ &= c \{(\dot{\eta}'_1(\theta) \dot{\eta}_1(\theta))(\dot{\eta}'_2(\theta) \dot{\eta}_2(\theta)) - (\dot{\eta}'_1(\theta) \dot{\eta}'_2(\theta))^2\}^{1/2} \end{aligned} \quad (14)$$

onde c é uma constante.

4. EXEMPLOS

4.1 - UM EXEMPLO NO CASO UNIPARAMÉTRICO

Considerar o modelo de regressão não-linear dado em (1) com $\eta_i(\theta) = e^{-\theta t_i}$, $i=1, 2, \dots, n$. Uma boa parametrização $\beta = g(\theta)$ é dada por $g'(\theta) = c(\dot{\eta}'(\theta) \dot{\eta}(\theta))^{1/2}$ que implica numa priori de Jeffreys localmente uniforme para β . Assim, precisamos resolver a equação diferencial,

$$g(\theta) = c \int \left(\sum_{i=1}^n t_i^2 e^{-2\theta t_i} \right)^{1/2} d\theta \quad (15)$$

Considerando $\sum_{i=1}^n t_i^2 e^{-2\theta t_i} \cong \int_0^{t_n} t^2 e^{-2\theta t} dt$ e um planejamento tal

que $t_n \uparrow \infty$, temos:

$$g(\theta) \cong c \int (4\theta^3)^{-1/2} d\theta = c\theta^{-1/2} \quad (16)$$

A parametrização $\beta = c\theta^{-1/2}$ também foi considerada por Hougaard, 1984 e por Bates e Watts, 1980. No artigo de Bates e Watts, foi visto que a parametrização que remove o efeito de curvatura é a solução da equação diferencial,

$$\frac{g''(\theta)}{g'(\theta)} = - \frac{\sum_{i=1}^n t_i^3 e^{-2\theta t_i}}{\sum_{i=1}^n t_i^2 e^{-2\theta t_i}} \quad (17)$$

Observamos que esta equação diferencial pode ser simplificada para a forma dada em (15).

Como ilustração numérica, considere os dados da Tabela 1.

TABELA 1.

i	1	2	3	4	5	6	7
t _i	1	2	3	4	5	6	30
y _i	0,90	0,80	0,70	0,70	0,60	0,50	0,05

Nas Figuras 1 e 2 temos os gráficos das funções de verossimilhança nas parametrizações θ e $\beta = -\theta^{-1/2}$. Observamos que apesar de todas as aproximações utilizadas, temos uma melhor aproximação à densidade normal na parametrização β .

Poderíamos usar outras aproximações para resolver a equação diferencial (15). Considerando e $e^{-2\theta t_i} \approx 1 - 2(\theta t_i) + 2(\theta t_i)^2$ e um planejamento tal

que $\sum_{i=1}^n t_i^3 = 0$, temos:

$$g(\theta) = \sqrt{a} \int \sqrt{1 + \frac{b}{a} \theta^2} d\theta \quad (18)$$

onde $a = \sum_{i=1}^n t_i^2$ e $b = 2 \sum_{i=1}^n t_i^4$.

Com a transformação $\sqrt{b/a} \theta = \text{tg } \lambda$, temos:

$$g(\theta) = \frac{a}{\sqrt{b}} \int \sec^3 \lambda d\lambda \quad (19)$$

Portanto, uma boa parametrização é dada por:

$$\beta = g(\theta) = \frac{a}{2\sqrt{b}} \left\{ \sqrt{\frac{b}{a}} \theta \left(1 + \frac{b}{a} \theta^2\right)^{1/2} + \ln \left[\left(1 + \frac{b}{a} \theta^2\right)^{1/2} + \sqrt{\frac{b}{a}} \theta \right] \right\} \quad (20)$$

Nas Figuras 3 e 4, temos os gráficos das funções de verossimilhança nas parametrizações θ e $\beta = 2,640(1+14\theta^2)^{1/2} + 0,71 \ln[(1+14\theta^2)^{1/2} + 3,74\theta]$ considerando o modelo não-linear $Y_i = e^{-\theta t_i} + \epsilon_i$, $i=1,2,\dots,n$ e os dados da Tabela 2.

TABELA 2.

i	1	2	3	4	5	6	7
t_i	-3	-2	-1	0	1	2	3
y_i	1,35	1,10	1,00	1,00	0,90	0,80	0,70

Claramente observamos melhor normalidade para a verossimilhança na parametrização β .

FIGURA 1: PARAMETRIZAÇÃO θ

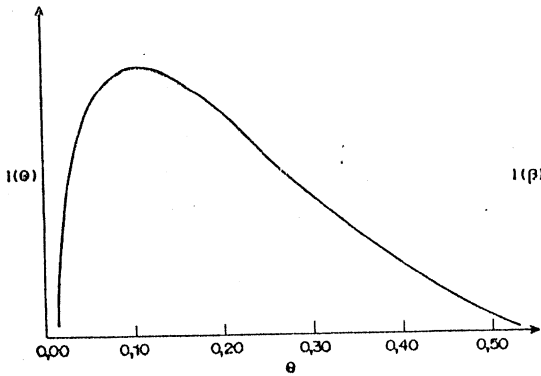


FIGURA 2: PARAMETRIZAÇÃO β

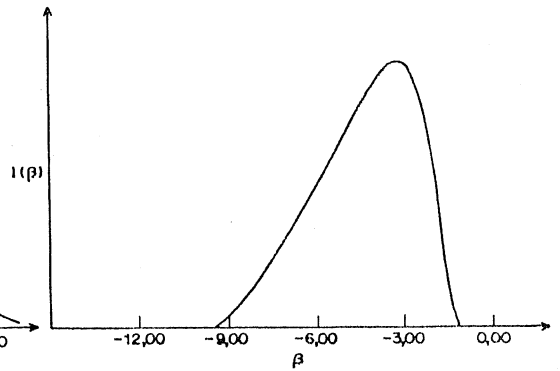


FIGURA 3: PARAMETRIZAÇÃO θ

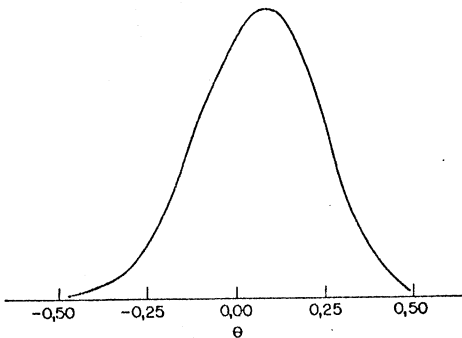
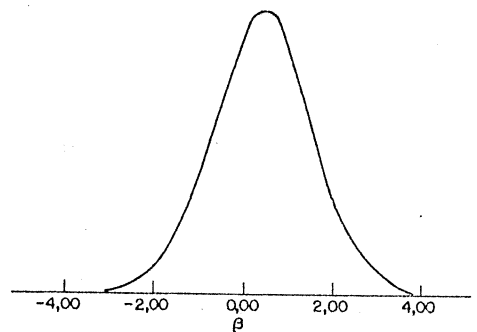


FIGURA 4: PARAMETRIZAÇÃO β



4.2 - UM EXEMPLO NO CASO MULTIPARAMÉTRICO

Supor o modelo não-linear $Y_i = \eta_i(\theta_1, \theta_2) + \epsilon_i$, $i=1, 2, \dots, n$ com $\eta_i(\theta_1, \theta_2) = f_i(\theta_1) + \theta_2$ e $f_i(\theta_1)$ é uma função não-linear de θ_1 . Observamos que

$$\dot{\eta}_1 = \left(\frac{df_1}{d\theta_1}, \dots, \frac{df_n}{d\theta_1} \right), \quad \dot{\eta}_2 = (1, 1, \dots, 1), \quad \dot{\eta}_1^T \dot{\eta}_1 = \sum_{i=1}^n \left(\frac{df_i}{d\theta_1} \right)^2, \quad \dot{\eta}_2^T \dot{\eta}_2 = n \text{ e } \dot{\eta}_1^T \dot{\eta}_2 = \sum_{i=1}^n \frac{df_i}{d\theta_1}.$$

Considerando $f_i(\theta_1) = e^{-\theta_1 t_i}$, temos $df_i(\theta_1)/d\theta_1 = -t_i e^{-\theta_1 t_i}$ e $(df_i/d\theta_1)^2 = t_i^2 e^{-2\theta_1 t_i}$. Supondo $g_2(\theta_1, \theta_2) = \theta_2$, precisamos encontrar $g_1(\theta_1, \theta_2)$ (de (14)) por:

$$g_1(\theta_1, \theta_2) = \left\{ n \sum_{i=1}^n t_i^2 e^{-2\theta_1 t_i} - \left(\sum_{i=1}^n t_i e^{-\theta_1 t_i} \right)^2 \right\}^{1/2} d\theta_1 \quad (21)$$

$$\text{Assumindo } \sum_{i=1}^n t_i^2 e^{-2\theta_1 t_i} \approx \int_0^{t_n} t^2 e^{-2\theta_1 t} dt, \quad \sum_{i=1}^n t_i e^{-\theta_1 t_i} \approx \int_0^{t_n} t e^{-\theta_1 t} dt$$

e considerando $t_n \uparrow \infty$, temos:

$$g_1(\theta_1, \theta_2) \approx \frac{1}{2} \int_0^\infty \theta_1^{-3/2} (n - 4\theta_1)^{1/2} d\theta_1 \quad (22)$$

Com a transformação $\theta_1 = n \cos^2 x/4$, obtemos,

$$\beta_1 = g_1(\theta_1, \theta_2) = 2 \left\{ \arccos \left(\frac{2\sqrt{\theta_1}}{n} \right) - \text{tg} \left[\arccos \left(\frac{2\sqrt{\theta_1}}{n} \right) \right] \right\} \quad (23)$$

Com uma ilustração numérica, considere os dados da Tabela 3 gerados a partir do modelo $E(Y_i) = e^{-\theta_1 t_i} + \theta_2$ com $\theta_1 = 0,50$ e $\theta_2 = 2,00$.

TABELA 3.

i	1	2	3	4	5	6	7	8	9	10
t_i	1	2	3	4	5	6	7	20	30	50
y_i	2,50	2,40	2,20	2,10	2,10	2,00	2,00	2,00	1,95	2,00

Nas Figuras 5 e 6 temos os contornos 0,95% para (θ_1, θ_2) e para (β_1, θ_2) onde β_1 é dado em (23). Observamos que na parametrização (β_1, θ_2) temos uma elipse mostrando a boa aproximação normal para a verossimilhança de β_1 e θ_2 .

FIGURA 5: PARAMETRIZAÇÃO (θ_1, θ_2)

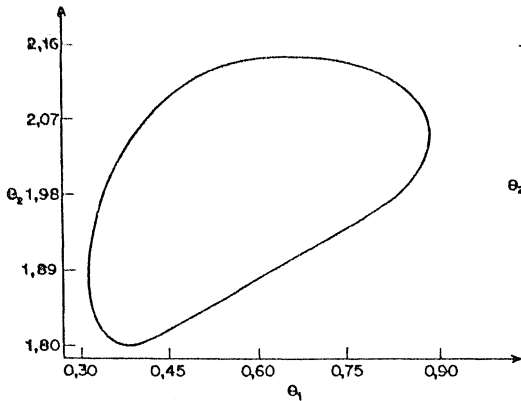
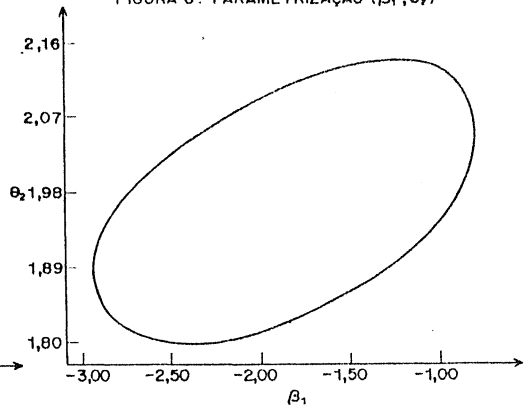


FIGURA 6: PARAMETRIZAÇÃO (β_1, θ_2)



REFERÊNCIAS

- ACHCAR, J.A.; SMITH, A.F.M. (1988). Aspects of reparametrization in approximate Bayesian inference (submetido para publicação).
- BATES, D.M.; WATTS, D.G. (1980). Relative curvature measures of nonlinearity. JRSS, B, 42(1), 1-25.
- BEALE, E.M.L. (1960). Confidence regions in non-linear estimation. JRSS, B, 22, 41-76.
- BOX, M.J. (1971). Bias in nonlinear estimation, JRSS, B, 33, 171-201.
- BOX, G.E.; TIAO, G.C. (1973). Bayesian Inference in statistical analysis. New York, Addison-Wesley.
- COX, D.R.; REID, N. (1987). Parameter orthogonality and approximate conditional inference. JRSS, B, 49(1), 1-39.
- DICICCIO, I.J. (1984). Parameter transformations and interval estimation, Biometrika, 71(3), 477-485.
- HOLLAND, P.W. (1973). Covariance stabilising transformations, Annals of Statistics, 1, 84-92.
- HOUGAARD, P. (1982). Parametrizations of non-linear models, JRSS, B, 44(2), 244-252.
- HOUGAARD, P. (1984). Parameter transformations in multiparameter regression models, Preprint 2, Copenhagen University.
- SPROTT, D.A. (1973). Normal likelihoods and their relation to large sample theory of estimation, Biometrika, 60, 457-465.
- SPROTT, D.A. (1980). Maximum likelihood in small samples: estimation in the presence of nuisance parameters, Biometrika, 67, 515-523.