Reparametrizações para inferência direta em modelos de regressão não linear

Walmes Marques Zeviani*

Resumo

Modelos de regressão não linear são considerados quando existe algum conhecimento preliminar sobre a relação entre variáveis. Tal conhecimento pode ser a respeito da própria natureza dos dados, uma equação diferencial e até mesmo a forma do diagrama de dispersão entre as variáveis. Em geral, seus parâmetros têm interpretação. Além disso, parâmetros de interesse, expressos como função dos parâmetros do modelo, são alvos de investigação. Para isso, o método delta, simulação Monte Carlo e procedimentos bootstrap são procedimentos adotados para fazer inferência. Além disso, uma reparametrização pode ser aplicada ao modelo de forma a representar esses parâmetros de interesse. Além de melhorar a interpretação do modelo, a presença do parâmetro alvo estende as possibilidades com relação a especificação de modelos e inferência estatística. O objetivo com esse trabalho é sistematizar o procedimento de aplicar reparametrizações. Ênfase é dada em modelos não lineares considerados em Ciências Agrárias. Uma lista com 17 modelos reparametrizados é fornecida. Breve discussão sobre os métodos de inferência é feita.

Palavras-chave: Função de parâmetros. Interpretação de parâmetros. Verossimilhança. Método delta. Curvatura.

1 INTRODUÇÃO

A ideia básica da regressão não linear é a mesma da regressão linear: relacionar uma resposta Y com um vetor de variáveis preditoras $\mathbf{x} = (x_1, \ldots, x_k)^\top$. Os modelos de regressão não linear são caracterizados pelo fato de a função de predição depender não linearmente de algum dos parâmetros. Embora não necessariamente, a regressão linear é usada para especificação de modelos puramente empíricos, enquanto que os modelos de regressão não linear são considerados quando existe algum conhecimento prévio para sustentar que a relação entre resposta e preditores segue uma particular forma funcional. Tal conhecimento pode ser desde uma equação diferencial que remete à particular modelo, como é o caso de modelos de crescimento, ou simplesmente uma restrição sobre a função, como o de a função ser monótona, típico de curvas de acúmulo, para a qual pode-se ter várias funções disponíveis.

*Doutorando em Estatística e Experimentação Agropecuária, DEX/UFLA. Professor do Departamento de Estatística - UFPR. Contato: walmes@ufpr.br.

Uma das principais vantagens do modelo de regressão não linear, é que frequentemente existe interpretação para a maioria de seus parâmetros (SCHABENBERGER; PIERCE, 2002). Esses parâmetros então passam ser o foco da investigação que, na sua forma mais simples, consiste em determinar intervalos de confiança e testar hipóteses. No entanto, uma situação comum é a necessidade de fazer inferência sobre uma função dos parâmetros (BENDER, 1996). Um exemplo simples é a equação de segundo grau $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, que é um modelo linear no qual o ponto crítico $x_c = -\beta_1/(2\beta_2)$ é alvo de inferência em situações de otimização de processos (BAS; BOYACI, 2007). Uma vez estimados seus parâmetros, inferência sobre x_c pode ser feita pelo método delta, por simulação Monte Carlo ou por métodos *bootstrap* (SEBER; WILD, 2003). Embora tais procedimentos permitam obter intervalos de confiança e conduzir testes de hipótese, existem ainda outras formas vantajosas de inferir ou modelar o parâmetro que são as extensões ligadas aos modelos não lineares de efeitos mistos (PINHEIRO; BATES, 2000) e a inferência bayesiana (DENISON et al., 2002).

2 REPARAMETRIZAÇÃO

Considere um modelo não linear

$$f(\mathbf{x}, \boldsymbol{\theta}) \tag{1}$$

em que f é uma função não linear que depende do vetor de covariáveis $\mathbf{x}$ e $\boldsymbol{\theta}$ é seu vetor de p parâmetros. Seja $\vartheta = g(\boldsymbol{\theta})$ o parâmetro de interesse em que g é uma função monótona e diferenciável em relação à $\boldsymbol{\theta}$. O objetivo com a reparametrização é fazer com que ϑ seja um elemento do vetor de parâmetros do modelo. Isso é obtido por substituição de algum dos p elementos de $\boldsymbol{\theta}$ por ϑ. Para isso, sistematizou-se o procedimento em três etapas:

1. Expressar o parâmetro de interesse como função dos elementos de $\boldsymbol{\theta}$, ou seja, $\vartheta = g(\boldsymbol{\theta})$;
2. Escolher um dos elementos θ_i de $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ para ser colocado em função de ϑ de tal forma a obter $\theta_i = h(\boldsymbol{\theta}_{-i}, \vartheta)$;
3. Substituir θ_i em (1) pela expressão obtida no passo anterior, $h(\boldsymbol{\theta}_{-i}, \vartheta)$, fazendo as simplificações convenientes. Assim o modelo (1) pode ser expresso como

$$f(\mathbf{x}, \boldsymbol{\theta}_{-i}, \vartheta)$$

.

A função h é a inversa de g em θ_i.

No passo 2 recomenda-se priorizar aquele elemento de $\boldsymbol{\theta}$ com menor

2.1 Reparametrização 1:1 - Modelos para acúmulo com ênfase na fração do total

O modelo Michaelis-Menten foi inicialmente proposto para para descrever a cinética de reações químicas (MICHAELIS; MENTEN, 1913). Tal modelo envolve uma função monótona crescente côncava a partir da origem. Atualmente observa-se a aplicação desse modelo em diversos contextos; um deles é a descrição do acúmulo de potássio liberado do solo (ZEVIANI et al., 2012). Sua forma funcional é

$$f(x; \theta_a, \theta_v) = \frac{\theta_a x}{\theta_v + x}, \qquad x \geq 0 \ (\mathrm{X}), \tag{2}$$

em que $\theta_a \geq 0$, é a assintota superior (Y) e representa o conteúdo total de nutriente liberado, e $\theta_v > 0$ é tempo de meia vida (X) ou tempo para fração meio.

2.2 Outros modelos

Além dos modelos considerados para exemplificar o procedimento de reparametrização, outros modelos frequentemente aplicados em Ciências Agrárias foram reparametrizados e estão apresentados nas tabelas de 1 à 3. A descrição de cada modelo, em termos de propriedades da função não linear, interpretação dos parâmetros é fornecida a seguir em uma lista numerada de acordo com as linhas das tabelas de 1 a 3.

Tabela 1: Reparametrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias

id	Modelo original	$\vartheta = g(\boldsymbol{\theta})$	$\theta_i = g^{-1}(\vartheta, \boldsymbol{\theta}_{-i})$	Modelo reparametrizado
1	$\frac{\theta_a x}{\theta_v + x}$	$\vartheta_q = \theta_v \left(\frac{q}{1-q}\right)$	$\theta_v = \vartheta_q \left(\frac{1-q}{q}\right)$	$\frac{\theta_a x}{\vartheta_q \left(\frac{1-q}{q}\right) + x}$
2	$\frac{\theta_a}{1 + \left(\frac{\theta_v}{x}\right)^{\theta_c}}$	$\vartheta_q = \theta_v \left(\frac{1-q}{q}\right)^{-1/\theta_c}$	$\theta_v = \vartheta_q \left(\frac{1-q}{q}\right)^{1/\theta_c}$	$\frac{\theta_a}{1 + \frac{1-q}{q} \left(\frac{\vartheta_q}{x}\right)^{\theta_c}}$
3	$\frac{\theta_a}{1 + \left(\frac{x}{\theta_v}\right)^{\theta_c}}$	$\vartheta_q = \theta_v \left(\frac{1-q}{q}\right)^{1/\theta_c}$	$\theta_v = \vartheta_q \left(\frac{1-q}{q}\right)^{-1/\theta_c}$	$\frac{\theta_a}{1 + \frac{1-q}{q} \left(\frac{x}{\vartheta_q}\right)^{\theta_c}}$
4	$\frac{\theta_a x^{\theta_c}}{\theta_v + x^{\theta_c}}$	$\vartheta_q = \left(\frac{\theta_v q}{1-q}\right)^{1/\theta_c}$	$\theta_v = \vartheta_q^{\theta_c} \frac{1-q}{q}$	$\frac{\theta_a x^{\theta_c}}{\vartheta_q \left(\frac{1-q}{q}\right) + x^{\theta_c}}$
5	$\theta_a(1 - \exp\{-\theta_c x\})$	$\vartheta_q = -\frac{\log(1-q)}{\theta_c}$	$\theta_c = -\frac{\log(1-q)}{\vartheta_q}$	$\theta_a(1 - \exp\{x \log(1-q)/\vartheta_q\})$
6	$\begin{cases} \theta_a(1 - \exp\{-\theta_1(x - \theta_0)\}) & , x \geq \theta_0 \\ 0 & , x < \theta_0 \end{cases}$	$\vartheta_q = \frac{\log(1-q)}{\theta_1} + \theta_0$	$\theta_1 = \frac{\log(1-q)}{\vartheta_q - \theta_0}$	$\theta_a \left(1 - \exp\left\{\log(1-q) \left(\frac{x - \theta_0}{\vartheta_q - \theta_0}\right)\right\}\right)$
7	$\theta_0 - \theta_1 x^{\theta_2}$	$\vartheta_q = \frac{q}{\theta_1}^{1/\theta_2}$	$\theta_2 = \frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}$	$\theta_0 - \theta_1 x^{\frac{\log(q) - \log(\theta_1)}{\log(\vartheta_q)}}$
8	$\theta_0 + \theta_1(1 - \theta_c^x)$	$\vartheta_q = \frac{\log(1 + q/\theta_1)}{\log(\theta_c)}$	$\theta_1 = -\frac{q}{1 - \theta_c^{\vartheta_q}}$	$\theta_0 - q \left(\frac{1 - \theta_c^x}{1 - \theta_c^{\vartheta_q}}\right)$

Tabela 2: (cont.) Reparametrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias.

id	Modelo original	$\vartheta = g(\boldsymbol{\theta})$	$\theta_i = g^{-1}(\vartheta, \boldsymbol{\theta}_{-i})$	Modelo reparametrizado
9	$\begin{cases} \theta_0 + \theta_1 x & , x \leq \theta_b \\ \theta_0 + \theta_1 \theta_b & , x > \theta_b \end{cases}$	$\vartheta_b = \theta_0 + \theta_1 \theta_b$	$\theta_0 = \vartheta_b - \theta_1 \theta_b$	$\begin{cases} \vartheta_b + \theta_1 (x - \theta_b) & , x \leq \theta_b \\ \vartheta_b & , x > \theta_b \end{cases}$
10	$\begin{cases} \theta_0 + \theta_1 x & , x \leq \theta_b \\ \theta_0 + \theta_1 \theta_b + \theta_2 (x - \theta_b) & , x > \theta_b \end{cases}$	$\vartheta_b = \theta_0 + \theta_1 \theta_b$	$\theta_0 = \vartheta_b - \theta_1 \theta_b$	$\begin{cases} \vartheta_b + \theta_1 (x - \theta_b) & , x \leq \theta_b \\ \vartheta_b + \theta_2 (x - \theta_b) & , x > \theta_b \end{cases}$
11	$\theta_0 + \theta_1 x + \theta_2 x^2$	$\vartheta_x = -\dfrac{\theta_1}{2\theta_2}$ $\vartheta_y = \theta_0 + \theta_1 \vartheta_x + \theta_2 \vartheta_x^2$	$\theta_1 = 2\theta_2 \vartheta_x$ $\theta_0 = \vartheta_y - \theta_1 \vartheta_x - \theta_2 \vartheta_x^2$	$\vartheta_y + \theta_2 (x - \vartheta_x)^2$
12	$\begin{cases} \theta_0 + \theta_1 x + \theta_2 x^2, \\ \quad x \leq -\theta_1/(2\theta_2) \\ \theta_0 + \theta_1 \left(\frac{-\theta_1}{2\theta_2}\right) + \theta_2 \left(\frac{-\theta_1}{2\theta_2}\right)^2, \\ \quad x > -\theta_1/(2\theta_2) \end{cases}$	$\vartheta_x = -\dfrac{\theta_1}{2\theta_2}$ $\vartheta_y = \theta_0 + \theta_1 \vartheta_x + \theta_2 \vartheta_x^2$	$\theta_1 = 2\theta_2 \vartheta_x$ $\theta_0 = \vartheta_y - \theta_1 \vartheta_x - \theta_2 \vartheta_x^2$	$\begin{cases} \vartheta_y + \theta_2 (x - \vartheta_x)^2 & , x \leq \vartheta_x \\ \vartheta_y & , x > \vartheta_x \end{cases}$
13	$x(\theta_0 + \theta_1 x)^{-1/\theta_2}$	$\vartheta_x = \dfrac{\theta_0}{\theta_1} \left(\dfrac{\theta_2}{1 - \theta_2}\right)$ $\vartheta_y = \vartheta_x \left(\dfrac{1 - \theta_2}{\theta_0}\right)^{1/\theta_2}$	$\theta_1 = \dfrac{\theta_0}{\vartheta_x} \left(\dfrac{\theta_2}{-1\theta_2}\right)$ $\theta_0 = (1 - \theta_2) \left(\dfrac{\vartheta_y}{\vartheta_x}\right)^{-\theta_2}$	$\vartheta_y \dfrac{x}{\vartheta_x} \left(1 - \theta_2 \left(1 - \dfrac{x}{\vartheta_x}\right)\right)^{-1/\theta_2}$
14	$\theta_0 x^{\theta_1} \exp\{-\theta_2 x\}$	$\vartheta_x = \theta_1/\theta_2$ $\vartheta_y = \theta_0 (\theta_1/\theta_2) \exp\{-\theta_1\}$ $\vartheta_p = \theta_2^{(\theta_1+1)}$	$\theta_1 = \theta_2 \vartheta_x$ $\theta_0 = \vartheta_y \left(\dfrac{1}{\vartheta_x}\right)^{\theta_1} \exp\{\theta_1\}$ $\theta_2 = \vartheta_p^{-1/(\theta_1+1)}$	$\vartheta_y \left(\dfrac{x}{\vartheta_x}\right)^{\dot{\theta}_1} \exp\{\dot{\theta}_1 (1 - x/\vartheta_x)\}$ $\dot{\theta}_1 \, : \, \dot{\theta}_1 - \vartheta_x \vartheta_p^{-1/(\dot{\theta}+1)}$

Tabela 3: (cont.) Reparametrizações desenvolvidas com ênfase na interpretação dos parâmetros de modelos de regressão não linear aplicados em Ciências Agrárias.

id	Modelo original	$\vartheta = g(\boldsymbol{\theta})$	$\theta_i = g^{-1}(\vartheta, \boldsymbol{\theta}_{-i})$	Modelo reparametrizado
15	$\dfrac{\theta_a}{1+\exp\{\theta_0+\theta_1 x\}}$	$\vartheta_q = \dfrac{1}{\theta_1}\left(\log\left(\dfrac{1-q}{q}\right)-\theta_0\right)$ $\vartheta_t = -\dfrac{\theta_1}{4}$	$\theta_0 = \log\left(\dfrac{1-q}{q}\right)-\theta_1\vartheta_q$ $\theta_1 = -4\vartheta_t$	$\dfrac{\theta_a}{1+\left(\frac{1-q}{q}\right)\exp\left\{-4\vartheta_t(x-\vartheta_q)\right\}}$
16	$\theta_a \exp\{-\exp\{\theta_0+\theta_1 x\}\}$	$\vartheta_q = \dfrac{\log(-\log(q))-\theta_0}{\theta_1}$	$\theta_1 = \dfrac{\log(-\log(q))-\theta_0}{\vartheta_x}$	$\theta_a \exp\{\log(q)\exp\{\theta_0(1-x/\vartheta_x)\}\}$
17	$\theta_r + \dfrac{\theta_s-\theta_r}{(1+\exp\{\theta_a+x\}^{\theta_n})^{\theta_m}}$	$\vartheta_i = -\theta_a - \log(\theta_m)/\theta_n$ $\vartheta_s = -\dfrac{\theta_n(\theta_s-\theta_r)}{(1-1/\theta_m)^{\theta_m+1}}$	$\theta_a = -\vartheta_i - \log(\theta_m)/\theta_n$ $\theta_s - \theta_r = -\frac{\vartheta_s}{\theta_n}(1+1/\theta_m)^{\theta_m+1}$	$\theta_r - \dfrac{\vartheta_s}{\theta_n}\dfrac{(1+1/\theta_m)^{\theta_m+1}}{(1+\exp\{\theta_n(x-\vartheta_i)\}/\theta_m)^{\theta_m}}$

3 ESTIMAÇÃO

Nessa seção será feita uma discussão sobre inferência sobre parâmetros em modelos de regressão não linear. A inferência baseada em verossimilhança será discutida, bem como inferência baseada na sua aproximação quadrática. Por fim, uma revisão do método delta será apresentada.

3.1 Verossimilhança

Seja $f(x, \boldsymbol{\theta})$ um modelo de regressão não linear considerado para descrever a média de uma variável aleatória Y. Considere que Y tenha distribuição normal com variância constante σ^2. Resumidamente, podemos escrever esse modelo como

$$\begin{aligned} Y &\sim \text{Normal}(\mu(x), \sigma^2) \\ \mu(x) &= f(x, \boldsymbol{\theta}). \end{aligned}$$

A função de verossimilhança do modelo é dada por

$$\text{L}(\boldsymbol{\theta}, \sigma^2) = \prod_{i=1}^{n} \phi(y_i, f(x_i, \boldsymbol{\theta}), \sigma^2), \tag{3}$$

em que ϕ representa a função densidade da distribuição Normal. O estimador de máxima verossimilhança são os valores $(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2)$ que tornam máximo o valor de L. Para estimação, é conveniente trabalhar com o logaritmo da função de verossimilhança

$$\ell(\boldsymbol{\theta}, \sigma^2) = \log \text{L}(\boldsymbol{\theta}, \sigma^2). \tag{4}$$

3.2 Método delta

O método delta é usado para aproximar a média e a variância de funções não lineares de variáveis aleatórias. Dentre suas aplicações, uma das mais comuns é relacionada à inferência sobre funções de parâmetros em modelos de regressão, como a razão entre parâmetros, transformação de um parâmetro, ou valor predito pelo modelo. Exemplos de funções de parâmetros estão na terceira coluna das tabelas 1 à 3.

4 CONSIDERAÇÕES FINAIS

A conclusão que se antecipa é que, uma vez que é possível reparametrizar o modelo para o parâmetro de interesse, inferência baseada na verossimilhança deve ser considerada, em segundo, sua aproximação quadrática, visto a capacidade de modelagem permitida por tais abordagens. Além do mais, as parametrizações devem ser avaliadas, seja por meio de medidas de curvatura, gráficos de perfil ou simulação, e deve ser escolhida àquela que tenha melhor compromisso entre propriedades estatísticas e de interpretação.

REFERÊNCIAS

BAS, D.; BOYACI, I. H. Modeling and optimization I: Usability of response surface methodology. **Journal of Food Engineering**, v. 78, n. 3, p. 836–845, fev. 2007.

BENDER, R. Calculating confidence intervals for summary measures of individual curves via nonlinear regression models. **International Journal of Bio-Medical Computing**, v. 41, n. 1, p. 13–18, mar. 1996.

DENISON, D. G. T. et al. **Bayesian Methods for Nonlinear Classification and Regression**. 1. ed. [S.l.]: Wiley, 2002. 296 p.

MICHAELIS, L.; MENTEN, M. Die kinetik der invertinwirkung. **Biochemische zeiturg**, v. 49, p. 333–369, fev. 1913.

PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-plus**. 1. ed. New York: Springer, 2000. 548 p.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil Sciences**. 1. ed. Boca Raton: CRC Press, 2002. 738 p.

SEBER, G. A. F.; WILD, C. J. **Nonlinear Regression**. Hoboken: Wiley, 2003. 752 p.

ZEVIANI, W. M. et al. Modelos não lineares para a liberação de potássio de estercos animais em latossolos. **Ciência Rural**, v. 42, n. 10, p. 1789–1796, out. 2012.