UM MODELO GERAL NÃO PARAMÉTRICO

Belmer Garcia Negrillo

RELATÓRIO TÉCNICO Nº 34/89

RESUMO – Neste trabalho generalizamos o modelo linear geral baseado em postos (GLM–R) apresentado por Hettmansperger and McKean (1977, 1983, 1984), onde sugerimos regras para a seleção dos escores, com a finalidade de obter estimadores mais eficientes e testes localmente mais poderosos

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
IMECC - UNICAMP
Caixa Postal 6065
13.081 - Campinas - SP
BRASIL

setembro – 1989

UM MODELO LINEAR GERAL NÃO PARAMETRICO

Belmer Garcia Negrillo
IMECC/UNICAMP

Resumo

Neste trabalho generalizamos o modelo linear geral baseado em postos (GLM-R) apresentado por Hettmansperger and McKean (1977, 1983, 1984), onde sugerimos regras para a seleção dos escores, com a finalidade de obter estimadores mais eficientes e testes localmente mais poderosos.

1. Introdução

Consideremos o modelo linear geral para o vetor de respostas Y, como uma função de α, β e da matriz de desenho X.

$$Y = [1\ X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon$$

com $Y_{N\times 1}$, $X_{N\times k}$, $\beta_{k\times 1}$, $\alpha_{1\times 1}$ e $\varepsilon_{N\times 1}$ vetor aleatorio com componentes i.i.d. com distribuição $F \in \Omega_0 = \{F\text{: F absolutamente continua e } F(0) = 1/2\}$.

Hettmansperger and McKean (1977, 1983, 1984), para obter o R-estimador de β, consideram a função distância de postos

$$D(Y - X\beta) = \sum_{i=1}^{N} a(R(Y_i - X_i'\beta))\ (Y_i - X_i'\beta)$$

onde $R(Y_i - X_i'\beta)$ é o posto de $Y_i - X_i'\beta$ em $Y_1 - X_1'\beta, \ldots, Y_N - X_n'\beta$ e $a(i)$ é uma função escore definida por

$$a(i) = -f'\left[F^{-1}\left(\frac{i}{N+1}\right)\right] \Big/ f\left[F^{-1}\left(\frac{i}{N+1}\right)\right] \qquad i = 1,\ldots,N$$

quando a função f de densidade de probabilidade dos erros é conhecida. Para o caso em que f é desconhecida, apresentamos na seção 6, algumas funções escores e regras para seleciona-los.

Para obter o R-estimador de α, consideramos a função distância de postos

$$D_1(\alpha,\hat{\beta}) = \sum_{i=1}^{N} b(R^+(Y_i - \alpha - X_i'\hat{\beta}))\, sgn(Y_i - \alpha - X_i'\hat{\beta})$$

onde

$R^+(Y_i-\alpha-X_i'\hat{\beta})$ e o posto de $|Y_i-\alpha-X_i'\hat{\beta}|$ em $|Y_1-\alpha-X_1'\hat{\beta}|,\dots,|Y_N-\alpha-X_N'\hat{\beta}|$.

$\hat{\beta}$ é o R-estimador de β e

$b(i) = a((i + N + 1)/2)$ para $i = 1,\dots,N$.

Com a finalidade de se estandarizar, os escores devem satisfazer

$E(a(.)) = 0$ $\qquad$ $Var(a(.))=1$

Por exemplo se $a^*(i)=i/(N+1)$ para $i=1,\dots,N$, são os escores de Wilcoxon, para obter os escores estandarizados temos que:

$E(a^*(i))= 1/2$ e $Var(a^*(i))=(1/12)N/(N+1)$

do que resulta

$$a(i) = (a^*(i)-E(a^*(i)))/(var(a^*(i)))^{1/2}$$
$$= (12)^{1/2}(i/(N+1)-(1/2))$$

2. R-estimador de β e α.

A função $D(Y - X\beta)$, definida na seção anterior, satisfaz a relação $D(Y - X\beta) = D(Y - 1\alpha - X\beta)$ para α escalar. Ademais é uma função não negativa, continua e convexa de β (Jaeckel (1972)), portanto, existem as derivadas parciais de $D(Y - X\beta)$.

O R-estimador de β é o valor $\hat{\beta}$ que minimiza $D(Y - X\beta)$. Assim, derivando em relação a β_j temos

$$\frac{\partial}{\partial \beta_j} D(Y - X\beta) = \sum_{i=1}^{N}(-x_{ij})\, a(R(Y_i - X_i'\beta)) = -\sum_{i=1}^{N} x_{ij}\, a(R(Y_i - X_i'\beta))$$

$$= -T_j(Y - X\beta) \qquad \text{para } j = 1,\dots,k\ ,$$

então, minimizar $D(Y - X\beta)$ é resolver o conjunto de equações não lineares

$$T_j(Y - X\beta) = 0 \qquad j = 1,\dots,k\ .$$

O R-estimador de α é o valor $\hat{\alpha}$ tal que $D_1(\hat{\alpha}, \hat{\beta}) = 0$ para $\hat{\beta}$ fixo.

Para determinar $\hat{\alpha}$, podemos usar um processo iterativo. Se $D_1(\alpha_r, \hat{\beta}) \neq 0$ escolhemos

$$\alpha_{r+1} > \alpha_r \quad \text{se} \quad D(\alpha_r, \hat{\beta}) > 0 \qquad \text{ou,}$$

$$\alpha_{r+1} < \alpha_r \quad \text{se} \quad D(\alpha_r, \hat{\beta}) < 0 \qquad \text{para} \quad r=1,2,\ldots$$

Se $D_1(\alpha,\hat{\beta}) = 0$ para $\alpha' < \alpha < \alpha''$ onde

$$\alpha'' = \sup \{\alpha : D_1(\alpha,\hat{\beta}) > 0\}$$

$$\alpha' = \inf \{\alpha : D_1(\alpha,\hat{\beta}) < 0\}$$

então

$$\hat{\alpha} = \frac{1}{2}(\alpha' + \alpha'')$$

3. Testes de Hipóteses para β

De Hettmansperger and McKean (1984), temos que para testar as hipóteses

$$H_0: H\beta = 0 \quad \text{vs} \quad H_1: H\beta \neq 0 \quad \text{com } H = (0, I_q)$$

a estatística do teste é dada por

$$F^* = \frac{(H\hat{\beta})' \, [H(X_c'X_c)^{-1}H']^{-1} (H\hat{\beta})}{\hat{\tau}^2}$$

onde $\hat{\beta}$ é o R-estimador de β, $X_c = X - 1\bar{X}$ e $\hat{\tau}$ é um estimador consistente de $\tau = (12^{1/2} \int f^2(x)dx)^{-1}$, sendo f a função de densidade dos erros.

A estatística F^* sob H_0, tem distribuição assintótica χ^2_q.

Na seção seguinte será apresentado dois métodos para estimar τ.

4. Estimação de τ

Existem principalmente duas classes de estimadores propostos para τ.

O primeiro, usando a função $D_1(.)$ utilizada para estimar α, proposto por Hettmansperger and McKean (1983) é dado por

$$\hat{\tau} = (N^{1/2}(U - L) / 2\, t_{\alpha/2}) \left[\frac{N}{N-k-1}\right]^{1/2}$$

onde U e L são as soluções C das equações

$$N^{1/2} \sum_{i=1}^{N} b(R^{+}(r_i - C))\, sgn(r_i - C) = t \quad \text{para}$$

$$t = t_{\alpha/2(N-k-1)} \qquad \text{e} \qquad t = -t_{\alpha/2(N-k-1)}$$

e r_i os resíduos.

A segunda classe de estimadores de τ é obtida estimando a função f(x), dada por

$$f_N(y) = \frac{1}{N h_N} \sum_{i=1}^{N} k \left[\frac{y - Y_i}{h_N}\right]$$

onde k(.) é uma função peso, não negativa, integrável e simétrica em 0 e h_N é uma sequência de constantes tal que $h_N \to 0$ e $Nh_N \to \infty$.

Para maior referência ver os trabalhos de Woodfield (1985); McDonald and Owen (1986).

5. Eficiência Relativa

A eficiência relativa entre os testes que utilizam o R-estimador e os que utilizam o estimador de mínimos quadrados é dado por

$$e(R,MQ) = \sigma^2 / \tau^2$$

Piere and Rauch (1984), compara os métodos GLM e GLMR para o modelo fatorial 3x4 com interação.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

usando a eficiência relativa assintótica de Pitman (ARE), para várias distribuições, cujos resultados são:

Distribuição	Normal	Normal contaminada	Exponencial dupla	Cauchy
ARE(GLMR,GLM)	0,9549	1,84	1,5	∞

6. Regras para a seleção da Função Escore a(.)

Nossa preocupação principal neste trabalho é a seleção da função escore, para obter estimadores mais eficientes e testes localmente mais poderosos, porque a seleção errada dos escores podem acarretar erros maiores que aqueles que cometeríamos utilizando o modelo linear geral (GLM), quando os erros não tem distribuição normal. Para isso utilizamos dois indicadores. Um indicador, para medir a simetria da distribuição das observações Y, para uma amostra $Y_1,\ldots,Y_N$ é dado por

$$\varphi_2 = [\bar{U}(0.05) - \bar{M}(0.5)] \,/\, [\bar{M}(0.5) - \bar{L}(0.05)]$$

onde $\bar{U}(\gamma)$, $\bar{M}(\gamma)$ e $\bar{L}(\gamma)$ são as médias aritméticas das $[N\gamma + 1]$ maiores, centrais e inferiores estatísticas de ordem de Y, respectivamente e [ ] significa parte inteira. O outro é um indicador da longitude da cauda da distribuição das observações Y, dada por Hogg (1972) e definido como

$$\varphi_1 = [\bar{U}(0.05) - \bar{L}(0.05)] \,/\, [\bar{U}(0.5) - \bar{L}(0.5)]$$

Para o caso de $c \geq 2$ amostras, para a i-ésima amostra determinamos φ_1 e φ_2 denotado por $\varphi_{1.i}$ e $\varphi_{2.i}$ respectivamente, para $i = 1,\ldots,c$ e determinamos

$$\bar{\varphi}_1 = \sum_{i=1}^{c} n_i \varphi_{1.i} / N , \qquad \bar{\varphi}_2 = \sum_{i=1}^{c} n_i \varphi_{2.i} / N \qquad N = \sum_{i=1}^{c} n_i$$

Sugerimos ainda regras para a seleção da função escore, para teste de locação. Estes escores satisfazem as restrições dos teoremas de simetria, normalidade e obtenção de testes localmente mais poderosos (Negrillo (1988)).

i) Se $1/2 \leq \bar{\varphi}_2 \leq 2$ e $\bar{\varphi}_1 < 2.24$ (simétrica e cauda curta) usar

$$a(i) = \begin{cases} (i - g - 1/2)/(N + 1) & \text{se } i < g \\ (i - N + g - 1/2)/(N + 1) & \text{se } i > N - g + 1 \\ 0 & \text{complemento} \end{cases}$$

onde $g = \left[\frac{N + 3}{4}\right]$

ii) Se $1/2 \leq \bar{\varphi}_2 \leq 2$ e $\bar{\varphi}_1 \in [2.24, 2.92]$ (simetrica e cauda média$^-$)

usar

$$a(i) = \Phi^{-1}\left(\frac{i}{N + 1}\right)$$ onde Φ é a distribuição $N(0,1)$

iii) Se $1/2 \leq \varphi_2 \leq 2$ e $\bar{\varphi}_1 \in (2.92, 3.8]$ (simétrica e cauda média $^+$)

usar

$$a(i) = \frac{i}{N + 1}$$

iv) Se $1/2 \leq \bar{\varphi}_2 \leq 2$ e $\bar{\varphi}_1 > 3.8$ (simétrica e cauda comprida)

usar

$$a(i) = \begin{cases} 1 & \text{se } i > \frac{1}{2}(N + 1) \\ 0 & \text{se } i \leq \frac{1}{2}(N + 1) \end{cases}$$

v) Se $\bar{\varphi}_2 > 2$ (assimetria à direita)

usar

$$a(i) = \begin{cases} i/(N+1) & \text{se } i \leq (N+1)/2 \\ \frac{1}{2} + \frac{1}{N+1} & \text{se } i > (N+1)/2 \end{cases}$$

vi) Se $\bar{\varphi}_2 < \frac{1}{2}$ (assimetria à esquerda)

usar

$$a(i) = \begin{cases} i/(N+1) & \text{se } i \geq (N+1)/2 \\ 1/2 & \text{se } i < (N+1)/2 \end{cases}$$

REFEFENCIAS

1. Hettmansperger, T. P. and McKean, J. W. (1977). "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Model"
Technometrics, Vol. 19, nº 3, pg 275-284

2. Hettmansperger, T. P. and McKean, J. W. (1983). "A Geometric Interpretation of Inference Based on Rank in the Linear Model" Journal of the American Statistical Association, Vol. 78, n$^{\underline{o}}$ 384, pg 885-893

3. Hettmansperger, T. P. (1984). "Statistical Inference Based on Ranks" John Wiley e Sons

4. Hogg, R. V. (1982). "On Adaptive Statistical Inference" Commum. Statist. Theor. Meth. 11 (22) pg 2531-2542

5. Jaeckel, L.A (1972). "Estimating regression Coefficientes by Minimizing the Dispersion of The Residuals" Annals of Mathematical Statistics, 43, 1449-1458

6. McDonald, J. A. and Owen, A. B. (1986). "Smoothing with Split Linear Fits" Technometrics Vol. 28 n$^{\underline{o}}$ 2, pg 195-208

7. Negrillo, B. G. (1988). "Metodos Não-Parametricos Uni e Multivariados" Relatório Interno IMECC/UNICAMP, a ser publicado

8. Pirie, W. R. and Rauch, H. L. (1984). "Simulated Efficiencies of Tests and Estimators from General Linear Models Aanalysis based on Ranks: The Two-way Layout with Interation" J. Statist. Comput. Simul., Vol. 20, pg 197-204

9. Woodfield, T. J. (1985). "The selection of Window Widths in Kernel Nonparametric Regression" J. Statist. Comput. Simul., Vol. 23, pg 113-122

RELATÓRIOS TÉCNICOS — 1989

01/89 — Uniform Approximation of Continuous Functions With Values in $[0,1]$ — *João B. Prolla.*

02/89 — On Some Nonlinear Iterative Relaxation Methods in Remote Sensing — *A. R. De Pierro.*

03/89 — A Parallel Iterative Method for Convex Programming with Quadratic Objective — *Alfredo N. Iusem* and *Alvaro R. De Pierro.*

04/89 — Fifth Force, Sixth Force, and all that: a Theoretical (Classical) Comment — *Erasmo Recami* and *Vilson Tonin–Zanchin.*

05/89 — An Application of Singer's Theorem to Homogeneous Polynomials — *Raymundo Alencar.*

06/89 — Summhammer's Experimental Test of the Non–Ergodic Interpretation of Quantum Mechanics — *Vincent Buonomano.*

07/89 — Privileged Reference Frames in General Relativity — *Waldyr A. Rodrigues Jr.* and *Mirian E. F. Scanavini.*

08/89 — On the Numerical Solution of Bound Constrained Optimization Problems — *Ana Friedlander* and *José Mario Martínez.*

09/89 — Dual Extremun Principles for the Heat Equation Solved by Finite Element Methods I — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

10/89 — Local Convergence Theory of Inexact Newton Methods Based on Structured Least Chance Updates — *José Mario Martínez*

11/89 — Real Spin–Clifford Bundle and the Spinor Structure of Space–Time — *Waldyr A. Rodrigues Jr.* and *Vera L. Figueiredo.*

12/89 — A Multiplier Theorem on Weightèd Orlicz Spaces — *B. Bordin* and *J. B. Garcia.*

13/89 — Dual Extremum Principles For The Heat Equation Solved By Finite Element Methods II — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

14/89 — Dirac and Maxwell Equations in the Clifford and Spin–Clifford Bundels — *W. A. Rodrigues Jr.* and *E. Capelas de Oliveira.*

15/89 — Formal Structures, The Concepts of Covariance Invariance, Equivalent Reference Frames, and the Principle of Relativity — *W. A. Rodrigues Jr., M. E. F. Scanavini* and *L. P. de Alcantara.*

16/89 — Local Minimizers of a Quadratic Function With a Spherical Constraint — *José Mario Martínez.*

17/89 — On Pseudo–Convex Polycircular Domains In Banach Spaces — *Mário C. Matos.*

18/89 — On Circular and Special Units of an Abelian Number Field — *Trajano Nóbrega.*

19/89 — Implementing Algorithms for Solving Sparse Nonlinear Systems of Equations — *Márcia A. Gomes–Ruggiero, José Mario Martínez* and *Antonio Carlos Moretti*.

20/89 — An Algorithm for Solving Nonlinear Least-Squares Problems with a New Curvilinear Search — *José Mario Martínez* and *Rita Filomena Santos*.

21/89 — Covariant Spinors, Algebraic Spinors, Operator Spinors and their Relationship — *V. L. Figueiredo, E. Capelas de Oliveira* and *W. A. Rodrigues Jr.*

22/89 — The Column–Updating Method for Solving Nonlinear Equations in Hilbert Space — *Márcia A. Gomes–Ruggiero* and *José Mario Martínez*.

23/89 — Automorphisms Control Systems and Observability — *Victor Ayala Bravo*

24/89 — Adjoint Functors Arise Everywhere — *T. M. Viswanathan*.

25/89 — New Sum Rules of Special Functions — *E. Capelas de Oliveira*.

26/89 — Minimal Realizations Under Controllability — *Luiz A. B. San Martin* and *Victor Ayala Bravo*.

27/89 — Optimization of Burg's Entropy Over Linear Constraints - *Yair Censor, Alvaro R. De Pierro* and *Alfredo N. Iusem*.

28/89 — Magnetic Monopoles without String in the Kähler–Clifford Algebra Bundle: A Geometrical Interpretation — *Adolfo Maia Jr., Erasmo Recami, Waldyr A. Rodrigues Jr.* and *Marcio A. F. Rosa*.

29/89 — On the Asymptotic Behavior of Some Alternate Smoothing Series Expansion Iterative Methods — *Alvaro R. De Pierro* and *Alfredo N. Iusem*.

30/89 — Tamanho da Amostra para Testes de Hipóteses Não Paramétricos — *Belmer Garcia Negrillo*.

31/89 — Unusual Black–Holes: About Some Stable (Nno–Evaporating) Extremal Solutions of Einstein Equations — *Vilson Tonin–Zanchin* and *Erasmo Recami*.

32/89 — Finite Form of Proper Orthocronous Lorentz Transformations and its Dynamical Interpretation — *J. Ricardo R. Zeni* and *Waldyr A. Rodrigues Jr.*

33/89 — Periodic Orbits Near the Boundary of a 3–Dimensional Manifold — *J. Sotomayor* and *M. A. Teixeira*.