

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

HIGOR DA CRUZ NEVES

Modelos Lineares Generalizados Semiparamétricos

Juiz de Fora
2021

HIGOR DA CRUZ NEVES

Modelos Lineares Generalizados Semiparamétricos

Monografia apresentada ao Programa de graduação em Estatística da Universidade Federal de Juiz de Fora como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Clécio da Silva Ferreira

Juiz de Fora

2021

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Neves, Higor.
Modelos Lineares Generalizados Semiparametricos / Higor
Neves. -- 2021.
40 f. : il.

Orientador: Clecio Ferreira
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2021.

1. Modelo Linear Generalizado Semiparamétrico. 2. Escore de Newton. 3. Spline cúbico. 4. Validação Cruzada. 5. Matriz de informação de Fisher. I. Ferreira, Clecio, orient. II. Título.

HIGOR DA CRUZ NEVES

Modelos Lineares Generalizados Semiparamétricos

Monografia apresentada ao Programa de graduação em Estatística da Universidade Federal de Juiz de Fora como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Aprovada em 17 de março de 2021.

BANCA EXAMINADORA

Professor Doutor Clécio da Silva Ferreira
Universidade Federal de Juiz de Fora

Professora Doutora Camila Borelli Zeller
Universidade Federal de Juiz de Fora

Professor Doutor Tiago Maia Magalhães
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Muito obrigado.

RESUMO

Nesta pesquisa, foi realizada a expansão do Modelo Linear Generalizado (MLG) para Modelo Linear Generalizado Semiparamétrico através da adição de uma curva não paramétrica na componente sistemática do MLG. Foram consideradas para o estudo as distribuições Binomial, Gamma, Inversa Gaussiana, Normal e Poisson, e suas principais ligações. Os nós da curva não paramétrica foram construídos através dos métodos de Eilers e Marks (1996) e de Green e Silverman (1994). A seleção do parâmetro de suavização foi feita por Validação Cruzada. Foram estudados os processos de estimação dos parâmetros e da curva não paramétrica através do método iterativo de Escore de Newton, bem como desenvolvidos os cálculos da função escore e matriz de informação de Fisher. Finalmente, com o auxílio do *software* estatístico R (R Core Team, 2021), foram elaborados os códigos de estimação e realizados estudos de simulação que avaliaram a eficiência do método de estimação dos parâmetros, além da aplicação em dados de pólen de Ambrosia de Kalamazoo em 1993.

Palavras-chave: Modelo Linear Generalizado Semiparamétrico. Escore de Newton. Spline cúbico. Validação Cruzada. Matriz de informação de Fisher.

ABSTRACT

In this research, the Generalized Linear Model (GLM) was extended to the Generalized Semiparametric Linear Model by adding a non-parametric curve to the MLG systematic component. The Binomial, Gamma, Gaussian Inverse, Normal and Poisson distributions and their main links were considered for the study. The nodes of the non-parametric curve were constructed using the methods of Eilers and Marks (1996) and Green and Silverman (1994). The selection of the smoothing parameter was made by Cross Validation. The methods for estimating the parameters and the non-parametric curve were developed using Newton's Scoring iterative method, as well as the calculations of the Fisher information matrix and the score function. Lastly, with the aid of the statistical software R (R Core Team, 2021), the estimation codes were elaborated and simulation studies were carried out to evaluate the efficiency of the parameter estimation method, in addition to the application in pollen data from Kalamazoo Ambrosia in 1993.

Keywords: Generalized Semiparametric Linear Model. Newton's Scoring. Cubic spline. Cross validation. Fisher information matrix.

LISTA DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1: Ajuste feito com um valor relativamente ideal para o parâmetro de suavização..... | 14 |
| Figura 2: Curva estimada com baixo valor de α | 15 |
| Figura 3: Curva estimada com alto valor de α | 15 |
| Figura 4: Gráficos da componente não paramétrica das 1000 replicações com curva não paramétrica cosseno..... | 34 |
| Figura 5: Gráficos da componente não paramétrica das 1000 replicações com curva não paramétrica do efeito doppler..... | 35 |
| Figura 6: Histograma da concentração de pólen em escala original (a), transformação raiz (b), e transformação raiz acrescido de 0,5 (c)..... | 36 |
| Figura 7: Gráficos da raiz da concentração e pólen por temperatura (a), velocidade do vento (b), dias da estação (c)..... | 37 |
| Figura 8: Curva não paramétrica estimada..... | 38 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1: Principais distribuições pertencentes à família exponencial | 20 |
| Tabela 2: Ligações canônicas das principais distribuições pertencentes à família exponencial | 21 |
| Tabela 3: Outras funções de ligação e suas respectivas distribuições. | 21 |
| Tabela 4: Média, Desvio Padrão das estimativas (SD) e Média dos Desvios Padrão dos Estimadores dos parâmetros (SDE) do modelo Gama com curva não paramétrica cosseno e nós selecionados pelo método de Eilers e Marx simulado com 1000 replicações e tempo em segundos para o cálculo da seleção de α e cálculo das estimativas..... | 34 |
| Tabela 5: Média, Desvio Padrão das estimativas (SD) e Média dos Desvios Padrão dos Estimadores dos parâmetros (SDE) do modelo Gama com curva não paramétrica do efeito doppler e nós selecionados pelo método de Eilers e Marx simulado com 1000 replicações, e tempo em segundos para o cálculo da seleção de α e cálculo das estimativas..... | 34 |
| Tabela 6: Estatísticas descritivas da raiz da concentração de pólen acrescida de 0,5 | 36 |
| Tabela 7: α e seu grau de liberdade ($GL\alpha$) e BIC para cada método de estimação aplicados . | 38 |
| Tabela 8: Estimativas calculados com o modelo Gama com ligação log..... | 38 |

SUMÁRIO

| | | |
|-------|--|----|
| 1 | INTRODUÇÃO | 9 |
| 1.1 | MODELO NÃO PARAMÉTRICO..... | 10 |
| 1.1.1 | Mínimos Quadrados Penalizados Modelo Não Paramétrico | 11 |
| 1.1.2 | Estimação da curva f | 11 |
| 1.1.3 | Parâmetro de Suavização α | 13 |
| 1.1.4 | Validação Cruzada..... | 16 |
| 1.2 | MODELO SEMIPARAMÉTRICO..... | 17 |
| 1.2.1 | Mínimos Quadrados Penalizados Modelo Semiparamétrico..... | 18 |
| 1.2.2 | Máxima Verossimilhança Penalizada | 18 |
| 1.3 | MODELO LINEAR GENERALIZADO | 20 |
| 1.3.1 | Função Desvio | 22 |
| 1.3.2 | Função Escore..... | 23 |
| 1.3.3 | Matriz Informação de Fisher..... | 24 |
| 1.3.4 | Estimação..... | 25 |
| 2 | MODELO LINEAR GENERALIZADO SEMIPARAMETRICO..... | 26 |
| 2.1 | FUNÇÃO ESCORE | 27 |
| 2.2 | MATRIZ DE INFORMAÇÃO..... | 28 |
| 2.3 | FUNÇÃO DESVIO | 29 |
| 2.3.1 | Normal | 30 |
| 2.3.2 | Poisson..... | 30 |
| 2.3.3 | Binomial..... | 30 |
| 2.3.4 | Gama..... | 30 |
| 2.3.5 | Normal Inversa | 31 |
| 2.4 | ESTIMAÇÃO DE β E f | 31 |
| 2.5 | ESTIMAÇÃO DE ϕ | 32 |
| 3 | APLICAÇÃO | 32 |
| 3.1 | ESTUDO DE SIMULAÇÃO | 32 |
| 3.2 | APLICAÇÃO EM DADOS REAIS..... | 35 |
| 4 | CONCLUSÕES..... | 39 |
| 5 | REFERÊNCIAS | 40 |

1 INTRODUÇÃO

Amplamente utilizado nas áreas de modelagem de dados, o Modelo de Regressão Linear Normal (MRLN) pode ser definido da forma:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n,$$

em que y_i representa o valor da i -ésima resposta observada; \mathbf{x}_i denota o vetor das p observações das variáveis explicativas; $\boldsymbol{\beta}$ é o vetor dos coeficientes de regressão e ϵ_i são os erros normais independentes. O MRLN trabalha sob quatro pressupostos: a relação linear entre as variáveis explicativas e a variável resposta, a independência entre os resíduos, com distribuição normal com média zero, e a homoscedasticidade. Para ser devidamente aplicado, o MRLN exige que todas suas pressuposições sejam atendidas, o que pode não ocorrer ao se trabalhar com dados reais.

Considerando os casos onde o pressuposto de linearidade entre uma das variáveis explicativas e a variável resposta não é atendido, os modelos semiparamétricos, segundo Ibacache *et al.* (2013), surgem como uma ferramenta poderosa em modelagem estatística devido à sua flexibilidade para modelar efeitos de variáveis explicativas que podem contribuir de forma paramétrica ou formas não paramétricas.

Definido por Green e Silverman (1994), o modelo assume que o relacionamento entre a variável resposta e as variáveis explicativas é representado da seguinte forma:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i) + \epsilon_i, \quad i = 1, \dots, n,$$

em que y_i denota a resposta do experimento, \mathbf{x}_i : denota o vetor de observações das variáveis explicativas, $\boldsymbol{\beta}$ é o vetor dos coeficientes de regressão, $f(t_i)$ é uma função de suavização, t_i são os valores da variável não paramétrica observada e ϵ_i são os erros aleatórios independentes.

Pensando agora nos casos onde o pressuposto de normalidade falha, por exemplo, para dados de contagem ou dados contínuos positivos, uma possível solução, proposta por Nelder e Wedderburn (1972), são os modelos lineares generalizados (MLGs), cujo conceito é de estender as possibilidades de distribuições que a variável resposta pode possuir, permitindo que esta pertença a qualquer distribuição da família exponencial. Os MLGs são definidos considerando Y_1, \dots, Y_n variáveis aleatórias independentes com distribuição de probabilidade:

$$f_Y(y_i; \theta_i, \phi) = \exp\left(\phi(y_i \theta_i - b(\theta_i)) + c(y_i, \phi)\right),$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas; $\phi > 0$ é denominado parâmetro de dispersão e θ é denominado parâmetro canônico; e pela componente sistemática:

$$g(\mu_i) = \eta_i,$$

em que $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é o preditor linear, $\mu_i = E(Y_i)$ e $g(\cdot)$ é a função de ligação.

Este trabalho tem como objetivo estudar uma possível solução para os casos onde tanto o pressuposto de normalidade quanto o de linearidade não são atendidos, denominado Modelo Linear Generalizado Semiparamétrico, ou MLGS, sendo obtida através da adição de uma curva suave na componente sistemática de um MLG. Para a seleção dos nós da curva será considerado os métodos de seleção de nós de Eilers e Marks (1996) e de Green e Silverman (1994); o parâmetro de suavização será obtido por meio de validação cruzada; a estimação dos parâmetros do modelo será feita com o auxílio da função escore e matriz de informação de Fisher, através do processo iterativo de escore de Newton. Por fim, para o aferimento da eficiência do método proposto, será realizado estudo de simulação e aplicação em dados reais.

1.1 MODELO NÃO PARAMÉTRICO

Podendo ser utilizado quando não é possível afirmar relação linear entre variável explicativa e variável resposta, o modelo de regressão não paramétrico é definido por Green e Silverman (1994) como:

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

onde $\mathbf{Y} = (y_1, \dots, y_n)^T$ representa o vetor das n respostas, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ os erros independentes, $f(t_i)$ função suave e $\mathbf{t} = (t_1, \dots, t_n)$ o vetor da variável explicativa observada.

Na estimação da curva $f(t_i)$ utiliza-se splines, que são curvas polinomiais obtidas com o auxílio de pontos denominados nós, e os coeficientes destes polinômios são estimados pelos métodos de mínimos quadrados ou máxima verossimilhança. A obtenção computacional destes nós pode ser feita através dos métodos de Green e Silverman (1994) ou Eilers e Marks (1996), onde o primeiro consiste em aplicar um nó em cada variação de t observada, e o segundo consiste em separar o intervalo de \mathbf{t} em $k > 0$ subintervalos de igual tamanho, aplicando os nós nos limites destes subintervalos.

1.1.1 Mínimos Quadrados Penalizados Modelo Não Paramétrico

Para se utilizar o método de mínimos quadrados para a estimação da curva $f(t)$, se faz necessário o acréscimo de um parâmetro que servirá como penalização do nível de ajuste, na suavidade ou rugosidade da curva.

Considerando que $f(t)$ seja duplamente diferenciável no intervalo $[a, b]$, uma maneira de se definir sua rugosidade é através do cálculo da integral de sua derivada segunda nesse intervalo:

$$J(f) = \int_a^b (f''(t))^2 dt$$

Definindo como parâmetro de suavização $\alpha > 0$, tem-se então que a soma de quadrados penalizados se dará da forma:

$$S(f) = \sum_{i=1}^n (y_i - f(t_i))^2 + \alpha \int_a^b (f''(t))^2 dt, \quad (1)$$

onde α é o parâmetro de suavização, e o estimador \hat{f} de mínimos quadrados penalizados é definido como o argumento mínimo de $S(f)$.

1.1.2 Estimação da curva f

Para a estimação da curva f de mínimos quadrados, se faz necessário o uso de splines, para este trabalho serão usados os splines cúbicos.

Dados números reais t_1, \dots, t_n , pertencentes a um intervalo $[a, b]$, da forma $a < t_1 < t_2 < \dots < t_{n-1} < t_n < b$, uma curva f definida nesse mesmo intervalo será considerada um spline cúbico se duas condições forem cumpridas: primeiramente, f será uma cúbica em cada um dos subintervalos $(a, t_1), (t_1, t_2), (t_2, t_3), \dots, (t_n, b)$; segundo, essas cúbicas se encontrarão em cada ponto t_i de forma suave, de modo que f será contínua e possuirá sua primeira e segunda derivada contínuas em todo intervalo $[a, b]$.

Estes pontos t_i são conhecidos como nós. Uma maneira de especificar um spline é usando quatro coeficientes polinomiais para cada cúbica, da forma:

$$f(t) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, \quad \text{para } t_i < t < t_{i+1}$$

onde a_i, b_i, c_i e d_i são constantes; $i = 0, \dots, n$; $t_0 = a$ e $t_{n+1} = b$. Para os casos onde o spline cúbico f no intervalo $[a, b]$ possui tanto segunda quanto terceira derivada iguais a 0 nos pontos a e b , ou seja, $c_0 = d_0 = c_n = d_n = 0$; dessa forma f será linear nos extremos e será conhecida como um spline cúbico natural (SCN), cuja vantagem será na redução da variância nos extremos do intervalo, onde a curva será aproximadamente linear.

Existe uma maneira mais conveniente de se especificar um spline cúbico natural, usando seu valor e o valor de sua segunda derivada no ponto t_i . Conhecido como Representação Via Valor da Derivada Segunda, será definido da forma:

$$f_i = f(t_i) \quad e \quad \gamma_i = f''(t_i), \quad i = 1, 2, \dots, n.$$

Pela definição de SCN, $\gamma_1 = \gamma_n = 0$. Dessa forma, obtém-se os vetores $\mathbf{f} = (f_1, \dots, f_n)$ e $\boldsymbol{\gamma} = (\gamma_2, \gamma_3, \dots, \gamma_{n-1})$, sendo possível definir uma fórmula em termos de \mathbf{f} e $\boldsymbol{\gamma}$ para o seu valor e de sua derivada no ponto t_i , dessa forma pode-se gerar gráficos de g para qualquer grau de precisão desejado.

Existem casos onde \mathbf{f} e $\boldsymbol{\gamma}$ podem não representar um SCN, dessa forma, Green e Silverman (1994) definem uma condição necessária para os vetores poderem representar um SCN dada sua sequência de nós. Tal condição depende de duas matrizes \mathbf{Q} e \mathbf{R} . Definindo $h_i = t_i + 1, i = 1, \dots, n$, \mathbf{Q} é definida como a matriz $n \times (n - 2)$ dos elementos $q_{ij}, i = 1, \dots, n$ e $j = 2, \dots, n - 1$ dados por:

$$q_{j-1,j} = h_{j-1}^{-1}, \quad q_{jj} = -h_{j-1}^{-1} - h_{j1}^{-1} \quad e \quad q_{j+1,j} = h_j^{-1}$$

e $q_{ij} = 0$ para $|i - j| \geq 2$ e as colunas de \mathbf{Q} são numeradas da mesma forma não padrão de $\boldsymbol{\gamma}$.

A matriz simétrica \mathbf{R} $(n - 2) \times (n - 2)$ possuirá tanto suas linhas como colunas numeradas da mesma forma não padrão de $\boldsymbol{\gamma}$, e seus elementos r_{ij} serão da forma:

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i), \quad i = 2, \dots, n - 1,$$

$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i, \quad i = 2, \dots, n - 2,$$

e $r_{ij} = 0$ para $|i - j| \geq 2$.

A matriz \mathbf{R} será diagonal dominante, ou seja, $|r_{ii}| > \sum_{i \neq j} |r_{ij}| \quad \forall i$, e também será positiva definida. Dessa forma é definida a matriz \mathbf{K} da forma:

$$\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T.$$

Os vetores \mathbf{f} e $\boldsymbol{\gamma}$ definirão um spline cúbico natural se a condição a seguir for satisfeita:

$$\mathbf{Q}^T \mathbf{f} = \mathbf{R}\boldsymbol{\gamma}$$

e se satisfeita, a penalidade da rugosidade irá satisfazer:

$$\int_a^b f''(t)^2 dt = \boldsymbol{\gamma}^T \mathbf{R} \boldsymbol{\gamma} = \mathbf{f}^T \mathbf{K} \mathbf{f}$$

e, conseqüentemente, pode-se reescrever a Equação (1) de soma de quadrados penalizados da forma:

$$S(\mathbf{f}) = (\mathbf{Y} - \mathbf{f})^T (\mathbf{Y} - \mathbf{f}) + \alpha \mathbf{f}^T \mathbf{K} \mathbf{f} \quad (2)$$

A unicidade do minimizador da curva spline pode ser obtido através da expansão da Equação (2):

$$\begin{aligned} S(\mathbf{f}) &= (\mathbf{Y} - \mathbf{f})^T (\mathbf{Y} - \mathbf{f}) + \alpha \mathbf{f}^T \mathbf{K} \mathbf{f} \\ &= \mathbf{f}^T (\mathbf{I} + \alpha \mathbf{K}) \mathbf{f} - 2\mathbf{Y}^T \mathbf{f} + \mathbf{Y}^T \mathbf{Y} \end{aligned} \quad (3)$$

Por $\alpha \mathbf{K}$ ser não-negativa definida, $\mathbf{I} + \alpha \mathbf{K}$ será estritamente positiva definida. Logo, $S(\mathbf{f})$ possuirá um mínimo único dado por:

$$\mathbf{f} = (\mathbf{I} + \alpha \mathbf{K})^{-1} \mathbf{Y}, \quad (4)$$

podendo ser verificado reescrevendo a Equação (3) da forma:

$$(\mathbf{f} - (\mathbf{I} + \alpha \mathbf{K})^{-1} \mathbf{Y})^T (\mathbf{I} + \alpha \mathbf{K}) (\mathbf{f} - (\mathbf{I} + \alpha \mathbf{K})^{-1} \mathbf{Y}) \quad (5)$$

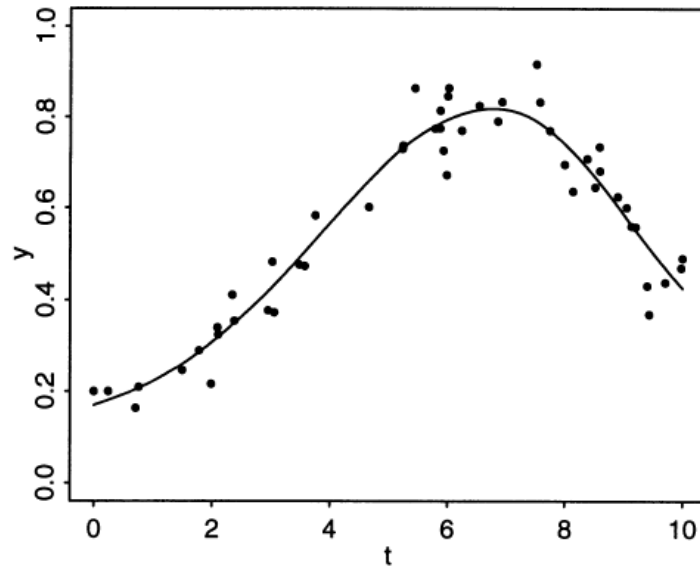
acrescido da constante que depende apenas de \mathbf{Y} . Como $\mathbf{I} + \alpha \mathbf{K}$ é estritamente positiva, a Equação (5) será estritamente positiva, exceto quando \mathbf{f} satisfaz (4), se tornando zero.

1.1.3 Parâmetro de Suavização α

Importante para a definição da rugosidade e na qualidade de ajuste da curva, o parâmetro de suavização α trabalha como um peso, definindo a importância do comprimento da curva f na minimização de $S(f)$.

Um valor de α ideal seria aquele onde a curva estimada fosse suave, porém ainda mantivesse a qualidade do ajuste, como pode ser visto na Figura 1.

Figura 1: Ajuste feito com um valor relativamente ideal para o parâmetro de suavização

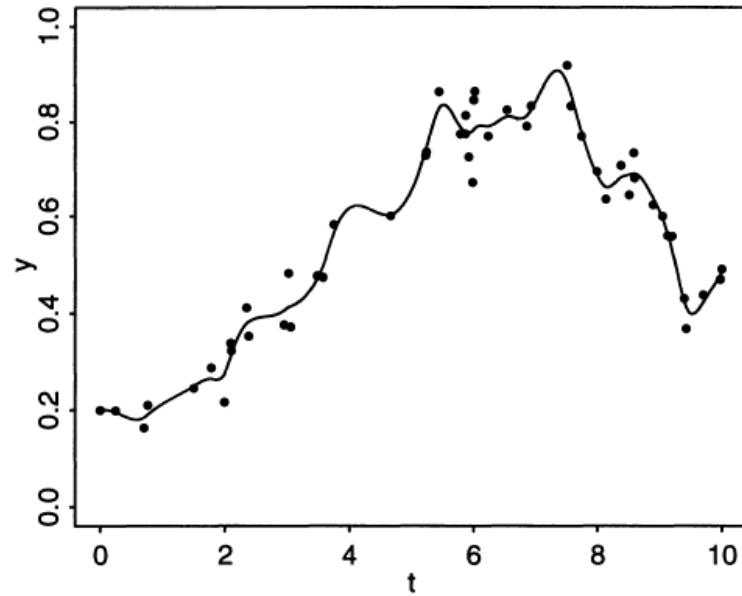


Fonte: Green e Silverman (1994).

Com base nos seus possíveis valores, existem dois extremos a serem mencionados, sendo eles conhecidos como *Overfitting* e *Underfitting*.

Se o valor de α tender a zero, o erro quadrático $\sum_{i=1}^n (y_i - f(t_i))^2$ de $S(f)$ terá um peso maior, e conseqüentemente, ao minimizar $S(f)$, teremos que $f(t_i) \rightarrow y_i \forall i$, ou seja, a curva tenderá a tocar todas as observações, causando o *Overfitting*, como pode se notar na Figura 2.

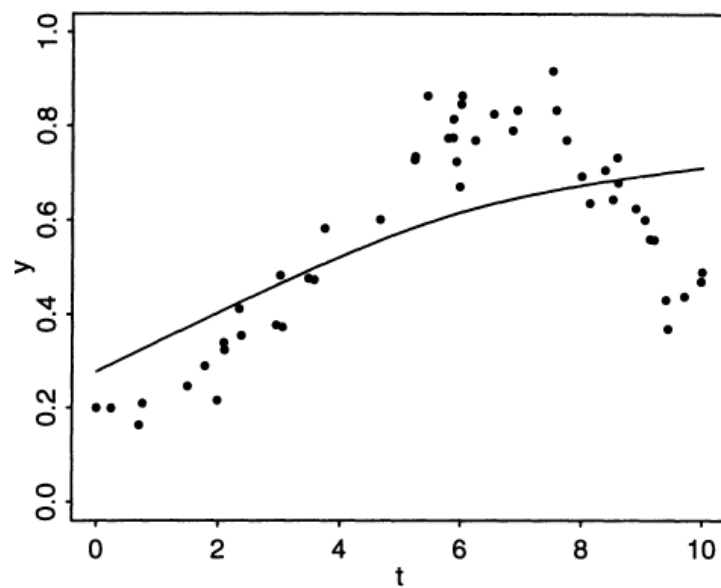
Figura 2: Curva estimada com baixo valor de α



Fonte: Green e Silverman (1994)

Se o valor de α tender a infinito, o comprimento da curva f , dado por $\int_a^b (f''(t))^2 dt$ da função $S(f)$ terá um peso maior, e conseqüentemente, ao se minimizar esta função, é possível observar que a curva tenderá à uma reta, uma vez que está sendo minimizando um arco com base em seu comprimento, causando o *Underfitting*, exemplificado na Figura 3.

Figura 3: Curva estimada com alto valor de α



Fonte: Green e Silverman (1994)

Na literatura existem diversos procedimentos automáticos distintos de se obter o melhor valor para o parâmetro de suavização. Bastante utilizada na literatura, por autores como Bowman (1984), Hardle e Marron (1985) e Stone (1974, 1977), a validação cruzada será utilizada neste estudo como método para a obtenção do parâmetro de suavização.

1.1.4 Validação Cruzada

Supondo um erro aleatório com média zero, a curva de regressão f tem a propriedade de que, se uma observação Y for tomada em um ponto t , o valor $f(t)$ é o melhor preditor de Y em termos de erro quadrático médio. Possuindo já fixado o valor do parâmetro de suavização α , e considerando Y_i em t_i como sendo nova observação, omitida com base no conjunto de dados utilizados para estimar a própria curva, denotando $\hat{f}(t_i, \alpha)$ a curva estimada para a i -ésima observação e $\hat{f}_{(-i)}(t_i, \alpha)$ a curva estimada para a i -ésima observação sem usar a mesma.

A eficácia do procedimento com o parâmetro α é dada pela função:

$$CV(\alpha) = n^{-1} \sum_{i=1}^n \left(Y_i - \hat{f}_{(-i)}(t_i, \alpha) \right)^2$$

Considerando:

$\mathbf{f} = \mathbf{H}(\alpha)\mathbf{Y}$, em que

$$\mathbf{H}(\alpha) = (\mathbf{I} + \alpha\mathbf{K})^{-1}$$

$\mathbf{H}(\alpha)$ é denominada matriz *hat* ou matriz de projeção, então é possível reescrever $CV(\alpha)$ como:

$$CV(\alpha) = n^{-1} \sum_{i=1}^n \left(\frac{Y_i - \hat{f}(t_i, \alpha)}{1 - H_{ii}} \right)^2$$

em que H_{ii} é o i -ésimo elemento da diagonal de \mathbf{H} . Para este trabalho será usada a validação cruzada generalizada, definida por Golub, Heath e Wahba (1979), dada por:

$$GCV(\alpha) = n^{-1} \frac{\sum_{i=1}^n (Y_i - \hat{f}(t_i, \alpha))^2}{(1 - n^{-1} \text{tr}(\mathbf{H}(\alpha)))^2}$$

em que $\text{tr}(\mathbf{H}(\alpha))$ é o traço da matriz $\mathbf{H}(\alpha)$.

O objetivo da validação cruzada generalizada é selecionar um valor de α que minimize o $GCV(\alpha)$. Como a função $GCV(\alpha)$ não garante um mínimo global, a minimização do $CV(\alpha)$ será feita nesse trabalho através da minimização para diversos valores iniciais ou pelo processo denominado *greed*.

A minimização para diversos valores iniciais consiste em, dado $\alpha \in [a, b]$, separa-se esse intervalo $[a, b]$ em $k > 0$ subintervalos igualmente espaçados e é feita minimização computacional do $GCV(\alpha)$ para cada subintervalo a partir dos seus devidos pontos médios s_k , ao final das k minimizações será selecionado o valor de α tal que $GCV(\alpha)$ tenha sido mínimo.

O processo *greed* consiste em, dado $\alpha \in [a, b]$, seleciona-se números reais s_1, \dots, s_k , pertencentes a este mesmo intervalo $[a, b]$, da forma $a = s_1 < s_2 < \dots < s_{k-1} < s_k = b$, de modo que $s_{i+1} - s_i = \frac{b-a}{k}$, $i = 1, \dots, k-1$, em outras palavras, são selecionados pontos igualmente distantes por todo o intervalo de α . O processo *greed* consiste em executar os cálculos das estimativas e GCV do modelo para todos os pontos s_i , e o α selecionado será o ponto s_i cujo GCV seja o mínimo. O processo *greed* tende a ser computacionalmente custoso, porém é extremamente robusto com a possibilidade da existência de mínimos locais.

1.2 MODELO SEMIPARAMÉTRICO

Considerando os bancos de dados onde as variáveis explicativas possuem tanto relação linear quanto não linear com a variável resposta, tanto o modelo paramétrico quanto o não paramétrico podem não fornecer as melhores estimativas por não considerarem parte da relação entre as variáveis. Nesses casos, é interessante o uso de um modelo que considere tanto a existência de uma componente paramétrica, quanto de uma componente não paramétrica, sendo um deles o modelo semiparamétrico. Definido por Green e Silverman (1994) como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + f(\mathbf{t}) + \boldsymbol{\varepsilon} \quad (6)$$

onde \mathbf{Y} representa o vetor das n respostas observadas, \mathbf{X} denota a matriz $n \times p$ das observações das variáveis explicativas; $\boldsymbol{\beta}$ o vetor dos p coeficientes de regressão, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ os erros independentes, com distribuição normal de média zero e variância σ^2 , ou seja, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$ e $f(\mathbf{t})$ uma curva não paramétrica.

1.2.1 Mínimos Quadrados Penalizados Modelo Semiparamétrico

Analogamente aos mínimos quadrados penalizados no modelo não paramétrico, se faz necessário o uso de uma penalização ao se usar mínimos quadrados no modelo semiparamétrico. Dessa forma, o estimador \hat{f} de mínimos quadrados penalizados para o modelo semiparamétrico será o argumento mínimo da soma de quadrados penalizados $S(f)$, dada por:

$$S(f) = \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \boldsymbol{\beta} - f(t_i))^2 + \alpha \int_a^b (f''(t))^2 dt$$

É possível reescrever $S(f)$ da forma:

$$S(f) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f}) + \alpha \mathbf{f}^T \mathbf{K} \mathbf{f} \quad (7)$$

em que a matriz \mathbf{N} , $n \times q$, é denominada matriz de incidência, composta por $N_{ij} = 1$ caso $t_i = s_j$ ou 0 caso contrário, e s_1, s_2, \dots, s_q os valores distintos ordenados do intervalo t_1, t_2, \dots, t_n . A Equação (7) terá seu mínimo quando $\boldsymbol{\beta}$ e \mathbf{f} satisfazerem:

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{N} \\ \mathbf{N}^T \mathbf{X} & \mathbf{N}^T \mathbf{X} + \alpha \mathbf{K} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{f} \end{pmatrix} = \begin{bmatrix} \mathbf{X}^T \\ \mathbf{N}^T \end{bmatrix} \mathbf{Y}$$

1.2.2 Máxima Verossimilhança Penalizada

Considere Y_1, Y_2, \dots, Y_n , ou seja, n observações independentes do modelo definido pela Equação (6), assim, $Y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{n}_i \mathbf{f}, \sigma^2)$, $i = 1, \dots, n$, em que \mathbf{n}_i é a i -ésima linha da matriz de incidência. \mathbf{N} O logaritmo da função de verossimilhança penalizada será dado da forma:

$$l_p(\boldsymbol{\beta}, \mathbf{f}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}$$

Note que $\frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}$ será o termo de penalização da log verossimilhança.

Para calcular os estimadores de máxima verossimilhança (EMV) teremos que derivar a log verossimilhança em função de seus respectivos parâmetros e igualar a derivada a zero.

EMV de $\boldsymbol{\beta}$:

$$\begin{aligned} \frac{\partial l_p(\boldsymbol{\beta}, \mathbf{f}, \sigma^2)}{\partial \boldsymbol{\beta}} &= -\frac{1}{\hat{\sigma}^2} \mathbf{X}^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{N}\hat{\mathbf{f}}) = 0 \\ \mathbf{X}^T \mathbf{Y} - \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}^T \mathbf{N}\hat{\mathbf{f}} &= 0 \\ \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^T (\mathbf{Y} - \mathbf{N}\hat{\mathbf{f}}) \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{Y} - \mathbf{N}\hat{\mathbf{f}}) \end{aligned}$$

Perceba que o termo $(\mathbf{Y} - \mathbf{N}\hat{\mathbf{f}})$ no EMV de $\boldsymbol{\beta}$ representa o resíduo não paramétrico explicado por $\hat{\boldsymbol{\beta}}$.

EMV de \mathbf{f} :

$$\begin{aligned} \frac{\partial l_p(\boldsymbol{\beta}, \mathbf{f}, \sigma^2)}{\partial \mathbf{f}} &= \frac{1}{\sigma^2} \mathbf{N}^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{N}\hat{\mathbf{f}}) - \hat{\sigma}^2 \alpha \hat{\mathbf{f}} \mathbf{K} = 0 \\ \mathbf{N}^T \mathbf{Y} - \mathbf{N}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{N}^T \mathbf{N}\hat{\mathbf{f}} - \hat{\sigma}^2 \alpha \hat{\mathbf{f}} \mathbf{K} &= \mathbf{0} \\ (\mathbf{N}^T \mathbf{N} + \alpha \mathbf{K}) \hat{\mathbf{f}} &= \mathbf{N}^T \mathbf{Y} - \mathbf{N}^T \mathbf{X}\hat{\boldsymbol{\beta}} \\ \hat{\mathbf{f}} &= (\mathbf{N}^T \mathbf{N} + \hat{\sigma}^2 \alpha \mathbf{K})^{-1} \mathbf{N}^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

Perceba também que o termo $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ no EMV de \mathbf{f} representa o resíduo paramétrico explicado pela curva f .

EMV de σ^2 :

$$\frac{\partial l_p(\beta, f, \sigma^2)}{\partial \sigma^2} = \frac{n}{2\hat{\sigma}^2} + \frac{1}{4\hat{\sigma}^4} (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{f})^T (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{f}) = 0$$

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{f})^T (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{f})}{n}$$

1.3 MODELO LINEAR GENERALIZADO

Propostos por Nelder e Wedderburn (1972), os Modelos Lineares Generalizados são definidos considerando Y_1, \dots, Y_n variáveis aleatórias independentes com distribuição de probabilidade pertencente à família exponencial, dada por:

$$f_Y(y_i; \theta_i, \phi) = \exp\left(\phi(y_i\theta_i - b(\theta_i)) + c(y_i, \phi)\right), \quad (8)$$

onde, $E(Y_i) = \mu_i = b'(\theta_i)$, $Var(Y_i) = \phi^{-1}V_i$, sendo $V_i = \frac{d\mu_i}{d\theta_i}$ função de variância e $\phi^{-1} > 0$ o parâmetro de dispersão; e a componente sistemática:

$$g(\mu_i) = \eta_i \quad (9)$$

onde $\eta_i = \mathbf{x}_i^T \beta$ é denominado preditor linear e $g(\cdot)$ é a função de ligação.

A Tabela 1 apresenta as principais distribuições pertencentes à família exponencial.

Tabela 1: Principais distribuições pertencentes à família exponencial

| Distribuição | $b(\theta)$ | θ | ϕ | $V(\mu)$ |
|----------------|----------------------|--|--------------------|----------------|
| Normal | $\frac{\theta^2}{2}$ | μ | σ^2 | 1 |
| Poisson | e^θ | $\log(\mu)$ | 1 | μ |
| Binomial | $\log(1 + e^\theta)$ | $\log\left(\frac{\mu}{1 - \mu}\right)$ | n | $\mu(1 - \mu)$ |
| Gama | $-\log(-\theta)$ | $-\frac{1}{\mu}$ | $\frac{1}{(CV)^2}$ | μ^2 |
| Normal Inversa | $-\sqrt{-2\theta}$ | $-\frac{1}{2\mu^2}$ | ϕ | μ^3 |

Fonte: Paula (2013).

Quando o parâmetro θ coincide com o preditor linear, em outras palavras, quando $\theta_i = \eta_i = g(\mu_i)$, a função de ligação $g(\cdot)$ é considerada uma ligação canônica. A Tabela 2 apresenta as ligações canônicas das principais distribuições.

Tabela 2: Ligações canônicas das principais distribuições pertencentes à família exponencial

| Distribuição | Normal | Binomial | Poisson | Gama | N. Inversa |
|--------------|--------------|---|--------------------|------------------------|--------------------------|
| Ligação | Identidade | Logit | Log | Inversa | Inversa Quadrada |
| | $\mu = \eta$ | $\log\left(\frac{\mu}{1-\mu}\right) = \eta$ | $\log(\mu) = \eta$ | $\frac{1}{\mu} = \eta$ | $\frac{1}{\mu^2} = \eta$ |

Fonte: Paula (2013).

Uma das vantagens de se fazer uso das ligações canônicas, segundo Paula (2013), é o fato de que as mesmas garantem a concavidade da função de verossimilhança de $\boldsymbol{\beta}$, e consequentemente facilita a obtenção de muitos resultados assintóticos; por exemplo, a concavidade da verossimilhança garante a unicidade da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$, quando essa existe. Para ligações não canônicas Wedderburn (1976) discute condições para a existência e da concavidade da verossimilhança de $\boldsymbol{\beta}$.

O objetivo da função de ligação é remover a restrição que μ_i possui de acordo com a distribuição. Dessa forma, ao se escolher uma $g(\cdot)$ tal que $g(\mu_i) \in \mathbb{R}$ evita-se a possibilidade de se estimar valores de $\boldsymbol{\beta}$ tais que valor de algum $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ não pertença a seu espaço paramétrico.

A Tabela 3 apresenta outras ligações também importantes.

Tabela 3: Outras funções de ligação e suas respectivas distribuições.

| | Ligação | Distribuição |
|----------------------------|---|---|
| <i>Cauchit</i> | $g(\mu_i) = F_Z(\mu_i),$ $Z \sim \text{Cauchy}(0,1)$ | Binomial |
| <i>Complemento log-log</i> | $g(\mu_i) = \log(-\log(1 - \mu_i))$ | Binomial |
| <i>Identidade</i> | $g(\mu_i) = \mu_i$ | Gama N. Inversa Normal Poisson |
| <i>Inversa</i> | $g(\mu_i) = \frac{1}{\mu_i}$ | Gama N. Inversa Normal |
| <i>Inversa quadrada</i> | $g(\mu_i) = \frac{1}{\mu_i^2}$ | N. Inversa |

| | | |
|---------------|--|---|
| <i>Log</i> | $g(\mu_i) = \ln(\mu_i)$ | Gama N. Inversa Normal Poisson |
| <i>Logit</i> | $g(\mu_i) = \ln\left(\frac{\mu_i}{1 - \mu_i}\right)$ | Binomial |
| <i>Probit</i> | $g(\mu_i) = \Phi(\mu_i)$ | Binomial |
| <i>Raiz</i> | $g(\mu_i) = \sqrt{\mu_i}$ | Poisson |

Fonte: Elaborada pelo autor (2021).

Outra distribuição importante pertencente à família exponencial é a distribuição Binomial Negativa, cujo modelo pode ser usado em dados de contagem com superdispersão como alternativa ao modelo de Poisson. Para maiores detalhes, ver Paula (2013), Seção 4.3.

1.3.1 Função Desvio

Definindo, sem perda de generalidade, o logaritmo da função de verossimilhança da forma:

$$L(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n L(\mu_i, y_i)$$

onde $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. No modelo saturado, onde $p = n$, $L(\boldsymbol{\mu}; \mathbf{y})$ será obtida por:

$$L(\mathbf{y}; \mathbf{y}) = \sum_{i=1}^n L(y_i, y_i)$$

Dessa forma, a estimativa de máxima verossimilhança de μ_i seria dada por $\tilde{\mu}_i = y_i$. Considerando $p < n$ a estimativa de $L(\boldsymbol{\mu}; \mathbf{y})$ é denotada por $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$, onde $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ e $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$.

Para o MLG, a função desvio será uma medida de distância entre a verossimilhança do ajuste do modelo saturado, onde o número de parâmetros é igual ao número de observações ($p = n$), e a do modelo sob observação, onde o número de parâmetros é menor que o número de observações ($p < n$), sendo dada da forma:

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2(L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y}))$$

Sua finalidade será avaliar a qualidade do ajuste do MLG. Note que um valor relativamente pequeno para a função desvio significa que a verossimilhança do modelo sob

observação, com um menor número de parâmetros, possui valor próximo à verossimilhança do ajuste do modelo saturado, ou seja, o modelo com um menor número de parâmetros possui um ajuste de qualidade similar à do modelo saturado. Alternativamente, pode-se escrever a função desvio da forma:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left(y_i (\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i) \right)$$

onde $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta(\tilde{\mu}_i)$ são, respectivamente, as estimativas de máxima verossimilhança de θ para o modelo com $p < n$ parâmetros e modelo saturado.

1.3.2 Função Escore

Denote $L(\boldsymbol{\beta}, \boldsymbol{\phi})$ o logaritmo da função de verossimilhança. Inicialmente, para obter a função escore para o parâmetro $\boldsymbol{\beta}$, deve-se calcular a derivada:

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\phi})}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) x_{ij} - \mu_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) x_{ij} \right\} \\ &= \sum_{i=1}^n \phi \left\{ \sqrt{\frac{w_i}{V_i}} (y_i - \mu_i) x_{ij} \right\} \end{aligned}$$

onde $w_i = V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right)^2$ é conhecido como função peso.

Escrevendo de forma matricial, temos:

$$\frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\phi})}{\partial \boldsymbol{\beta}} = \boldsymbol{\phi} \mathbf{X}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu})$$

onde $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ é a matriz de pesos.

Concluindo assim, a função escore para $\boldsymbol{\beta}$ como:

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \phi) = \left[\frac{\partial L(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} \right]$$

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \phi) = \phi \mathbf{X}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu})$$

Note que no caso das ligações canônicas, $V(\boldsymbol{\mu}) = \mathbf{W}$, então, $\mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} = \mathbf{W}^{\frac{1}{2}} \mathbf{W}^{-\frac{1}{2}} = \mathbf{I}$, sendo assim a função escore de $\boldsymbol{\beta}$, nesses casos, será escrita como:

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \phi) = \phi [\mathbf{X}^T (\mathbf{Y} - \boldsymbol{\mu})]$$

A função escore de ϕ será dada por

$$\begin{aligned} \mathbf{U}_\phi(\boldsymbol{\beta}, \phi) &= \frac{\partial L(\boldsymbol{\beta}, \phi)}{\partial \phi} \\ &= \sum_{i=1}^n \{(y_i \theta_i - b(\theta_i))\} + \sum_{i=1}^n \frac{dc(y_i, \phi)}{d\phi} \end{aligned}$$

1.3.3 Matriz Informação de Fisher

A matriz informação de Fisher para $\boldsymbol{\beta}$ será definida da forma:

$$\begin{aligned} \mathbf{M}_{\beta\beta}(\boldsymbol{\beta}, \phi) &= -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \phi; \mathbf{Y})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right) \\ &= \phi \mathbf{X}^T \mathbf{W} \mathbf{X} \end{aligned}$$

De forma análoga a $\boldsymbol{\beta}$, matriz de informação de Fisher para ϕ será definida da forma

$$\begin{aligned} \mathbf{M}_{\phi\phi}(\boldsymbol{\beta}, \phi) &= -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \phi)}{\partial \phi^2} \right) \\ &= - \sum_{i=1}^n E \left(\frac{d^2 c(Y_i, \phi)}{d\phi^2} \right) \end{aligned}$$

É possível observar a ortogonalidade entre $\boldsymbol{\beta}$ e ϕ utilizando

$$\begin{aligned}
\mathbf{M}_{\beta\phi}(\boldsymbol{\beta}, \phi) &= -E\left(\frac{\partial^2 L(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \phi}\right) \\
\mathbf{M}_{\beta\phi}(\boldsymbol{\beta}, \phi) &= -E\left(\sum_{i=1}^n \sqrt{w_i V_i^{-1}} (Y_i - \mu_i) x_i\right) \\
\mathbf{M}_{\beta\phi}(\boldsymbol{\beta}, \phi) &= 0
\end{aligned} \tag{10}$$

1.3.4 Estimação

Defina \mathbf{U}'_{β} como a primeira derivada de \mathbf{U}_{β} em relação a $\boldsymbol{\beta}^T$. A estimação de $\boldsymbol{\beta}$ é feita através do processo iterativo de Newton Raphson, definida a partir da expansão da função escore de $\boldsymbol{\beta}$ sob um valor inicial $\boldsymbol{\beta}^{(0)}$, da forma:

$$\mathbf{U}_{\beta} \cong \mathbf{U}_{\beta}^{(0)} + \mathbf{U}'_{\beta}{}^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})$$

onde $\mathbf{U}_{\beta}^{(0)}$ e $\mathbf{U}'_{\beta}{}^{(0)}$ representam, respectivamente, \mathbf{U}_{β} e \mathbf{U}'_{β} avaliados em $\boldsymbol{\beta}^{(0)}$. Dessa forma, na iteração $k = 0, 1, \dots$, a estimativa de $\boldsymbol{\beta}$ se dará da forma:

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left((- \mathbf{U}'_{\beta})^{-1}\right)^{(k)} \mathbf{U}_{\beta}^{(k)}$$

Como $- \mathbf{U}'_{\beta}$ pode não ser positiva definida, é comum fazer a aplicação do método de escore de Fisher, através de sua substituição por seu valor esperado $E(- \mathbf{U}'_{\beta}) = -E(\mathbf{U}'_{\beta}) = \mathbf{M}_{\beta\beta}$, que por sua vez é a matriz de informação de Fisher, resultando no processo iterativo:

$$\begin{aligned}
\boldsymbol{\beta}^{(k+1)} &= \boldsymbol{\beta}^{(k)} + (\mathbf{M}_{\beta\beta}^{-1})^{(k)} \mathbf{U}_{\beta}^{(k)} \\
&= (\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{z}^{(k)}
\end{aligned}$$

$k = 0, 1, \dots$, onde $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})$. É possível observar pela Equação (10) que a estimação de $\boldsymbol{\beta}$ independe de ϕ , ou seja, como já comentado, os estimadores são ortogonais.

Para obter a estimação de ϕ , inicialmente iguala-se a função escore U_ϕ a zero, obtendo:

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(\mathbf{y}; \hat{\mu}) - \sum_{i=1}^n (y_i \tilde{\theta}_i - b(\tilde{\theta}_i))$$

em que $D(\mathbf{y}; \hat{\mu})$ é o desvio do modelo sob investigação. Por depender do desvio, de $b(\cdot)$ e $\tilde{\theta}$, o estimador de máxima verossimilhança de ϕ dependerá da distribuição usada.

Para as distribuições Normal e Normal Inversa, a estimativa de máxima verossimilhança de ϕ será dada por:

$$\hat{\phi} = \frac{n}{D(\mathbf{y}; \hat{\mu})}$$

Na distribuição Gama, a estimativa de máxima verossimilhança de ϕ será obtida resolvendo a equação:

$$2n (\log(\hat{\phi}) - \psi(\hat{\phi})) = D(\mathbf{y}; \hat{\mu})$$

onde ψ representa a função digama dada por $\psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$, onde $\Gamma(\phi)$ é a função gama dada por $\Gamma(\phi) = \int_0^\infty t^\phi e^{-t} dt$.

Baseado na estatística de Pearson, sob a suposição de que $\hat{\beta}$ tenha sido consistentemente estimado, um outro estimador consistente para ϕ , será dado por:

$$\hat{\phi} = \frac{(n-p)}{\sum_{i=1}^n \left(\frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right)} \quad (11)$$

2 MODELO LINEAR GENERALIZADO SEMIPARAMETRICO

Podendo agir como uma solução para os casos onde os pressupostos de normalidade e linearidade entre uma das variáveis explicativas e a variável resposta não são atendidos, o

Modelo Linear Generalizado Semiparametrico, ou MLGS, se dará pela adição de uma componente não paramétrica ao preditor linear, de modo que:

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i)$$

onde $f(t_i) = \mathbf{n}_i^T \mathbf{f}$, e \mathbf{n}_i é o vetor da i -ésima coluna da matriz de incidência N .

Analogamente à verossimilhança penalizada do modelo semiparamétrico, a verossimilhança penalizada para o MLGS será dada pela adição de uma penalização à verossimilhança do MLG, da forma:

$$L(\boldsymbol{\beta}, \mathbf{f}) = \sum_{i=1}^n \phi(y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c(y_i, \phi) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}$$

2.1 FUNÇÃO ESCORE

Considere $L(\boldsymbol{\beta}, \mathbf{f}, \phi)$ ou $L(\boldsymbol{\beta}, \mathbf{f})$ a função de verossimilhança penalizada. Inicialmente, para obter a função escore, deve-se calcular as derivadas:

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) x_{ij} - \mu_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) x_{ij} \right\} \\ &= \sum_{i=1}^n \phi \left\{ \sqrt{\frac{w_i}{V_i}} (y_i - \mu_i) x_{ij} \right\} \end{aligned}$$

Escrevendo de forma matricial:

$$\frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \boldsymbol{\beta}} = \boldsymbol{\phi} \mathbf{X}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu})$$

Derivando agora em relação a f :

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \mathbf{f}} &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{df_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{df_j} \right\} - \frac{\alpha}{2} \frac{\partial \mathbf{f}^T \mathbf{K} \mathbf{f}}{\partial \mathbf{f}} \\
&= \sum_{i=1}^n \phi \left\{ y_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) \mathbf{n}_i - \mu_i V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right) \mathbf{n}_i \right\} - \alpha \mathbf{K} \mathbf{f} \\
&= \sum_{i=1}^n \phi \left\{ \sqrt{\frac{w_i}{V_i}} (y_i - \mu_i) \mathbf{n}_i \right\} - \alpha \mathbf{K} \mathbf{f}
\end{aligned}$$

Escrevendo na forma matricial:

$$\frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \mathbf{f}} = \phi \mathbf{N}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu}) - \alpha \mathbf{K} \mathbf{f}$$

Concluindo assim, a função escore como:

$$\mathbf{U}(\boldsymbol{\beta}, \mathbf{f}) = \begin{bmatrix} \frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \boldsymbol{\beta}} \\ \frac{\partial L(\boldsymbol{\beta}, \mathbf{f})}{\partial \mathbf{f}} \end{bmatrix}$$

$$\mathbf{U}(\boldsymbol{\beta}, \mathbf{f}) = \phi \begin{bmatrix} \mathbf{X}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu}) \\ \mathbf{N}^T \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \boldsymbol{\mu}) - \alpha \mathbf{K} \mathbf{f} \end{bmatrix} \quad (12)$$

Note que no caso das ligações canônicas, $V(\mu) = W$, então, $\mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} = \mathbf{W}^{\frac{1}{2}} \mathbf{W}^{-\frac{1}{2}} = \mathbf{I}$.

Sendo assim, para esses casos, a função escore será escrita como:

$$\mathbf{U}(\boldsymbol{\beta}, \mathbf{f}) = \phi \begin{bmatrix} \mathbf{X}^T (\mathbf{Y} - \boldsymbol{\mu}) \\ \mathbf{N}^T (\mathbf{Y} - \boldsymbol{\mu}) - \alpha \mathbf{K} \mathbf{f} \end{bmatrix}$$

2.2 MATRIZ DE INFORMAÇÃO

A matriz de informação esperada de Fisher será definida da forma:

$$\mathbf{M}(\boldsymbol{\beta}, \mathbf{f}) = -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \mathbf{f})}{\partial(\boldsymbol{\beta}, \mathbf{f}) \partial(\boldsymbol{\beta}, \mathbf{f})^T} \right),$$

tal que

$$\mathbf{M}(\boldsymbol{\beta}, \mathbf{f}) = \begin{bmatrix} I_{\beta\beta} & I_{\beta f} \\ I_{f\beta} & I_{ff} \end{bmatrix} \quad (13)$$

em que:

$$I_{\beta\beta} = -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \mathbf{f})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right) = \boldsymbol{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X}$$

$$I_{\beta f} = -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \mathbf{f})}{\partial \boldsymbol{\beta} \partial \mathbf{f}} \right) = \boldsymbol{\phi} \mathbf{X}^T \mathbf{W} \mathbf{N}$$

$$I_{f\beta} = -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \mathbf{f})}{\partial \mathbf{f} \partial \boldsymbol{\beta}} \right) = \boldsymbol{\phi} \mathbf{N}^T \mathbf{W} \mathbf{X}$$

$$I_{ff} = -E \left(\frac{\partial^2 L(\boldsymbol{\beta}, \mathbf{f})}{\partial \mathbf{f}^T \partial \mathbf{f}} \right) = \boldsymbol{\phi} \mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \mathbf{f}$$

Pela matriz esperada de Fisher é possível obter o erro padrão dos estimadores, dado por:

$$D(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}}) = \left(\text{diag} \left(\mathbf{M}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})^{-1} \right) \right)^{\frac{1}{2}}, \quad (14)$$

onde $\text{diag} \left(\mathbf{M}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})^{-1} \right)$ é o vetor da diagonal principal de $\mathbf{M}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})^{-1}$.

Assim como em MLG, o parâmetro $\boldsymbol{\phi}$ é ortogonal aos parâmetros de locação, neste caso a $\boldsymbol{\beta}$ e \mathbf{f} . Ou seja, $I_{\beta\phi} = I_{f\phi} = \mathbf{0}$.

2.3 FUNÇÃO DESVIO

Todo o cálculo para a função desvio é feito de forma similar ao MLG, como visto na sessão 1.3.1, obtendo a função desvio da forma:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left(y_i (\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i) \right) \quad (15)$$

em que $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta(\tilde{\mu}_i)$ são, respectivamente, as estimativas de máxima verossimilhança de θ para o modelo com $p < n$ parâmetros e modelo saturado. A única diferença entre a função desvio do MLG para a função desvio do MLGS será nos valores de $\hat{\theta}_i$

e $\tilde{\theta}_i$, devido ao fato que no MLGS tem-se $\boldsymbol{\mu}_i = g^{-1}(\boldsymbol{\eta}_i) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{n}_i \mathbf{f})$, em que \mathbf{n}_i é a i -ésima linha da matriz de incidência \mathbf{N} . A seguir será apresentada a função desvio para alguns casos particulares.

2.3.1 Normal

Para o modelo normal, $\theta_i = \mu_i$. Logo $\hat{\theta}_i = \hat{\mu}_i$ e $\tilde{\theta}_i = y_i$, nesse caso o desvio será dado por:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left(y_i(y_i - \hat{\mu}_i) + \frac{\hat{\mu}_i^2}{2} - \frac{y_i^2}{2} \right) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2,$$

que coincide com a soma de quadrado de resíduos.

2.3.2 Poisson

Para o modelo poisson, $\theta_i = \log(\mu_i)$. Logo $\hat{\theta}_i = \log(\hat{\mu}_i)$ e $\tilde{\theta}_i = \log(y_i)$. Dessa forma o desvio será dado por:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left(y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right)$$

2.3.3 Binomial

Assumindo $Y_i \sim \text{Binomial}(n_i, \mu_i)$, $i = 1, \dots, n$, tem-se que $\hat{\theta}_i = \log \left(\frac{\hat{\mu}_i}{n_i - \hat{\mu}_i} \right)$ e $\tilde{\theta}_i = \log \left(\frac{y_i}{n_i - y_i} \right)$ para $0 < y_i < n_i$, nesse caso o desvio será dado por:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left(y_i \log \left(\frac{y_i \hat{\mu}_i}{n_i} \right) + (n_i - y_i) \log \left(\frac{1 - \frac{y_i}{n_i}}{1 - \hat{\mu}_i} \right) \right)$$

2.3.4 Gama

Para o modelo gama, $\theta_i = \frac{1}{\mu_i}$. Logo $\hat{\theta}_i = \frac{1}{\hat{\mu}_i}$ e $\tilde{\theta}_i = \frac{1}{y_i}$, nesse caso o desvio, caso todos os valores sejam positivos, será dado por:

$$D(\mathbf{y}; \hat{\mu}) = 2 \sum_{i=1}^n \left(-\log\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right)$$

Note que caso alguma observação y_i seja igual a zero, o desvio será indeterminado, nesses casos, Mc-Cullagh e Nelder (1989) sugerem substituir o desvio por:

$$D^*(\mathbf{y}; \hat{\mu}) = 2\phi C(\mathbf{y}) + 2\phi \sum_{i=1}^n \left(\log(\hat{\mu}_i) + \frac{y_i}{\hat{\mu}_i} \right)$$

onde $C(\mathbf{y})$ é uma função arbitrária limitada, por exemplo $C(\mathbf{y}) = \sum_{i=1}^n \left(\frac{y_i}{1+y_i} \right)$.

2.3.5 Normal Inversa

Para a normal inversa, $\theta_i = -\frac{1}{2\mu_i^2}$, logo $\hat{\theta}_i = -\frac{1}{2\hat{\mu}_i^2}$ e $\tilde{\theta}_i = -\frac{1}{2y_i^2}$, dessa forma, o desvio será dado por:

$$D(\mathbf{y}; \hat{\mu}) = 2 \sum_{i=1}^n \left(\frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2} \right)$$

2.4 ESTIMAÇÃO DE $\boldsymbol{\beta}$ E \mathbf{f}

Considere $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{f})$, $\mathbf{U}(\boldsymbol{\tau})$ a matriz de escore e $\mathbf{K}(\boldsymbol{\tau})$ a matriz de informação de Fisher, e α conhecido.

Similar ao MLG, não é conhecida uma solução analítica para a estimativa dos parâmetros do modelo, portanto, será usado o processo iterativo de Newton-Raphson, descrito na sessão 1.3.4, definido a partir da expansão da função escore de $\boldsymbol{\beta}$ e \mathbf{f} sob um valor inicial $\boldsymbol{\tau}^{(0)} = (\boldsymbol{\beta}^{(0)}, \mathbf{f}^{(0)})$, obtendo, por fim, na iteração $k = 0, 1, \dots$, o processo iterativo:

$$\boldsymbol{\tau}^{(k+1)} = \boldsymbol{\tau}^{(k)} + \mathbf{M}^{-1}(\boldsymbol{\tau}^{(k)})\mathbf{U}(\boldsymbol{\tau}^{(k)}) \quad (16)$$

2.5 ESTIMAÇÃO DE ϕ

Similar ao MLG, a função escore de ϕ será dada por:

$$\begin{aligned} U_{\phi}(\boldsymbol{\beta}, \mathbf{f}, \phi) &= \frac{\partial L(\boldsymbol{\beta}, \mathbf{f}, \phi)}{\partial \phi} \\ &= \sum_{i=1}^n \{(y_i \theta_i - b(\theta_i))\} + \sum_{i=1}^n \frac{dc(y_i, \phi)}{d\phi} \end{aligned}$$

Note que tanto para o MLG quanto para o MLGS, a função escore de ϕ será a mesma. Consequentemente, fórmulas para sua estimação serão iguais, já descritas na sessão 1.3.4.

3 APLICAÇÃO

Para este trabalho foi desenvolvido no software estatístico *R Core Team (2021)* o algoritmo de estimação de MLG Semiparamétrico. O algoritmo foi escrito considerando as distribuições e ligações mencionadas nas Tabela 2 e Tabela 3, além dos métodos de seleção dos nós de Green e Silverman, e Eilers e Marx. Dado um valor do parâmetro de suavização α , faz-se uso da Função Escore (Equação (12)), e Matriz de Informação (Equação (13)), para a estimação dos parâmetros através do processo iterativo de Escore de Newton (Equação (16)), além dos cálculos dos desvios padrão dos estimadores (Equação (14)), e função desvio (Equação (15)). O processo também conta com a otimização do parâmetro de suavização α que minimiza o GCV. Os códigos estão disponíveis na plataforma *github* pelo *link*: <http://www.github.com/ClecioFerreira/SGLM> e *ad request* pelo e-mail: higor.cruz@ice.ufjf.br.

3.1 ESTUDO DE SIMULAÇÃO

Com o auxílio do software estatístico R, foram simuladas 1000 amostras de tamanho 50, 100 e 250 para a família Gamma com ligação log. Durante a simulação, para evitar a possibilidade de mínimos locais na função GCV, a obtenção do parâmetro de suavização α será feita através do processo *greed*.

O equipamento usado para a execução das simulações foi um notebook Dell® modelo Inspiron 5458, processador Intel® Core i5-5200U, 4 núcleos, sendo utilizado apenas um para as simulações, frequência de 2.20GHz e 3MB de cache; um único módulo de memória RAM DDR3 1600MHz de 8GB, sistema operacional Windows® 10 64 bits e SSD 250GB.

Considere o modelo:

$$g(\boldsymbol{\mu}) = \ln(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Nf} = \boldsymbol{\eta}$$

onde

$$\boldsymbol{\beta} = (\beta_1, \beta_2) = (4, -1);$$

$$x_{1i} \sim \text{Uniforme}(0,1) \text{ e } x_{2i} \sim \text{Uniforme}(0,1), i = 1, \dots, n;$$

\mathbf{X} matriz $n \times 2$ formada por $[x_1, x_2]$;

$$n = 50, 100, 250;$$

\mathbf{t}_1 um vetor composto por uma sequência de $\frac{n}{2}$ valores nos intervalos de $(-\frac{3}{2}\pi, \frac{5}{2}\pi)$ para cosseno e $(\frac{2}{3}\pi, \frac{5}{3}\pi)$ para efeito doppler;

Efeito doppler dado por: $doppler(t) = \cos(4\pi t) \exp\left(-\frac{t^2}{2}\right)$;

\mathbf{t} um vetor de tamanho n composto pelos valores do vetor $(\mathbf{t}_1, \mathbf{t}_1)$, ou seja, os valores de t_1 repetidos em sequência.

$$f(t) = \cos(t) \text{ e } doppler(t)$$

$$\boldsymbol{\mu} = \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Nf});$$

$$Y_i \sim \text{Gamma}\left(\phi, \frac{\phi}{\mu_i}\right), i = 1, \dots, n;$$

$$\phi = 100;$$

$$\alpha \in [0.1, 3];$$

Serão considerados 15 nós tanto para a curva cosseno quanto para o efeito doppler pelo método de Eilers e Marx.

Foram feitas também simulações considerando o método de seleção de nós de Green e Silverman, seus resultados foram menos satisfatórios em relação ao ajuste da curva e não serão introduzidos nessa monografia.

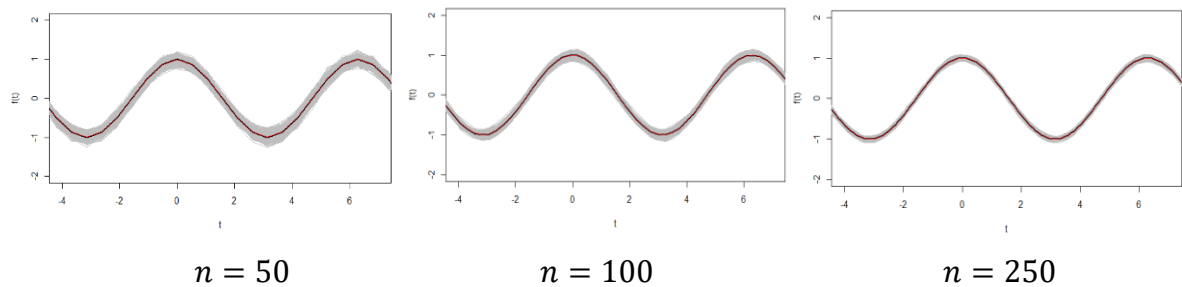
Dessa forma foram obtidas as estimativas apresentadas nas Tabela 4 e Tabela 5. Nas Figura 4 e Figura 5 são apresentadas as curvas estimadas em cinza, curva verdadeira em preto e médias das curvas estimadas em vermelho.

Tabela 4: Média, Desvio Padrão das estimativas (SD) e Média dos Desvios Padrão dos Estimadores dos parâmetros (SDE) do modelo Gama com curva não paramétrica cosseno e nós selecionados pelo método de Eilers e Marx simulado com 1000 replicações e tempo em segundos para o cálculo da seleção de α e cálculo das estimativas

| Parâmetro | Valor Real | $n = 50$ | | | $n = 100$ | | | $n = 250$ | | |
|--------------|------------|----------|--------|-------|-----------|--------|-------|-----------|-------|-------|
| | | Média | SD | SDE | Média | SD | SDE | Média | SD | SDE |
| β_1 | 4,000 | 3,994 | 0,056 | 0,046 | 3,999 | 0,037 | 0,034 | 4,000 | 0,021 | 0,021 |
| β_2 | -1,000 | -0,999 | 0,060 | 0,050 | -1,000 | 0,040 | 0,038 | -0,999 | 0,023 | 0,022 |
| ϕ | 100,000 | 158,767 | 43,981 | | 123,559 | 19,915 | | 108,398 | 9,954 | |
| Tempo | - | 0,350 | 0,123 | | 1,219 | 0,088 | | 3,689 | 1,676 | |
| α | - | 1,700 | 1,278 | | 1,873 | 1,344 | | 1,680 | 1,403 | |

Fonte: Elaborada pelo autor (2021).

Figura 4: Gráficos da componente não paramétrica das 1000 replicações com curva não paramétrica cosseno.



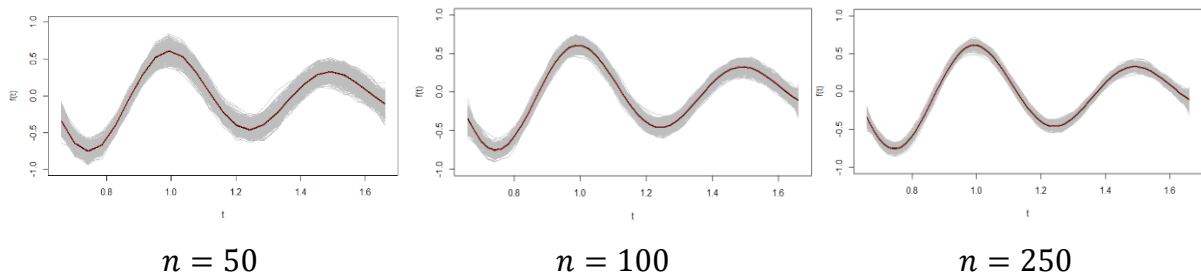
Fonte: Elaborada pelo autor (2021).

Tabela 5: Média, Desvio Padrão das estimativas (SD) e Média dos Desvios Padrão dos Estimadores dos parâmetros (SDE) do modelo Gama com curva não paramétrica do efeito doppler e nós selecionados pelo método de Eilers e Marx simulado com 1000 replicações, e tempo em segundos para o cálculo da seleção de α e cálculo das estimativas

| Parâmetro | Valor Real | $n = 50$ | | | $n = 100$ | | | $n = 250$ | | |
|--------------|------------|----------|--------|-------|-----------|--------|-------|-----------|--------|-------|
| | | Média | SD | SDE | Média | SD | SDE | Média | SD | SDE |
| β_1 | 4,000 | 4,002 | 0,062 | 0,054 | 3,997 | 0,039 | 0,035 | 4,000 | 0,023 | 0,022 |
| β_2 | -1,000 | -1,005 | 0,059 | 0,053 | -0,998 | 0,035 | 0,033 | -1,000 | 0,023 | 0,022 |
| ϕ | 100,000 | 157,170 | 42,344 | | 123,552 | 20,336 | | 108,365 | 10,397 | |
| Tempo | - | 0,316 | 0,089 | | 1,386 | 0,579 | | 1,834 | 0,909 | |
| α | - | 1,876 | 1,362 | | 1,842 | 1,340 | | 1,730 | 1,355 | |

Fonte: Elaborada pelo autor (2021).

Figura 5: Gráficos da componente não paramétrica das 1000 replicações com curva não paramétrica do efeito doppler



Fonte: Elaborada pelo autor (2021).

O objetivo dessa simulação foi aferir a validade dos cálculos propostos para função escore, matriz de informação de Fisher, validação cruzada e a estimação dos parâmetros do MLGS pelo processo iterativo de escore de Fisher. É possível notar a qualidade do ajuste dos parâmetros e da curva. As componentes paramétricas tiveram ajuste preciso para todos os valores de n tanto para a simulação com a curva cosseno quanto para a curva do efeito doppler. Para a curva cosseno os ajustes foram consideravelmente precisos para amostra de tamanho 50, e, como esperado, tiveram ajustes ainda mais precisos para $n = 100$ e 250, como pode ser observado na

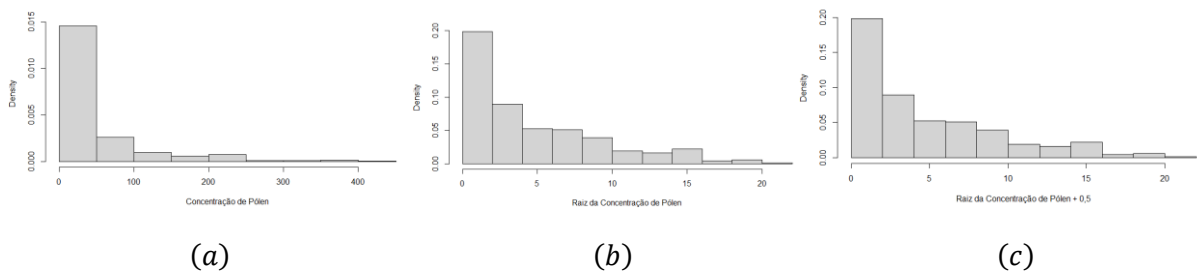
Figura 4. Observando o ajuste para o efeito doppler, por ser uma curva relativamente mais complexa que o cosseno, os ajustes para $n = 50$ possuem uma variação ligeiramente maior, com melhora considerável para maiores tamanho de amostra. Interessante notar que os valores médios de α selecionados não sofreram grandes alterações entre os tamanhos e amostra, indicando que α depende apenas da complexidade da curva, e independe do tamanho da amostra.

3.2 APLICAÇÃO EM DADOS REAIS

De acordo com Stark *et al.* (1997), alergias relacionadas ao pólen são uma doença comum, resultando em rinite alérgica e asma em cerca de 10% da população. Os dados que serão analisados foram coletados diariamente em durante a estação da Ambrosia no ano de 1991 a 1994, em Kalamazoo, Michigan, totalizando 335 observações. Como em Ruppert *et al.* o interesse é explicar a concentração de pólen de ambrosia diária, em $\text{gãos}/\text{m}^3$. A Figura 6 (a) representa o histograma da concentração de pólen em escala original, no qual é possível notar uma assimetria positiva com uma longa cauda que pode ser difícil de se ajustar sem o uso de

modelos para caudas pesadas. A Figura 6 (b) representa uma transformação raiz, que parece ter um comportamento ajustável pelos modelos propostos por este trabalho, onde será acrescido a essa transformação o valor de 0,5 para evitar possíveis erros nas estimações dos modelos gama e inversa gaussiana. A Figura 6 (c) e Tabela 6 representam, respectivamente, o histograma e as estatísticas descritivas da distribuição da concentração de pólen, sob a transformação raiz acrescida de 0,5.

Figura 6: Histograma da concentração de pólen em escala original (a), transformação raiz (b), e transformação raiz acrescida de 0,5 (c)



(a) Fonte: Elaborada pelo autor (2021).

(b)

(c)

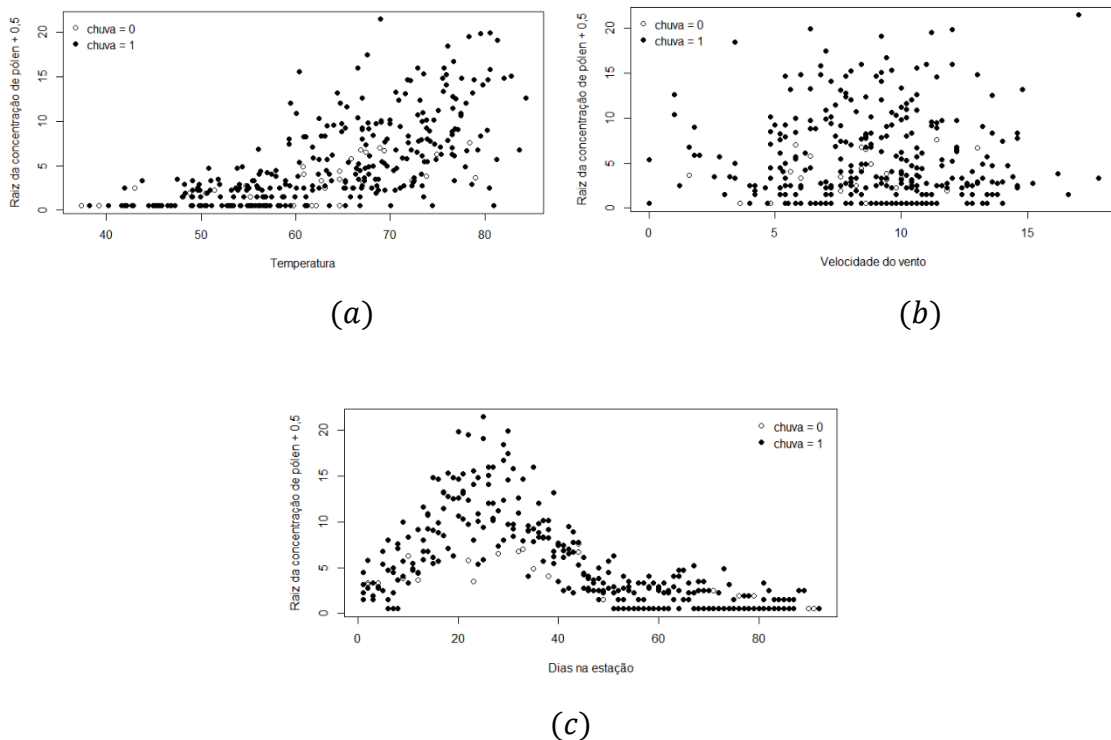
Tabela 6: Estatísticas descritivas da raiz da concentração de pólen acrescida de 0,5

| Mínimo | 1° quartil | Mediana | Média | 3° quartil | Máximo |
|--------------|------------|---------|-------|------------|--------|
| 0,500 | 1,500 | 3,500 | 5,214 | 7,983 | 21,476 |

Fonte: Elaborada pelo autor (2021).

Como o banco de dados também apresenta alguns dados ambientais, se mostra razoável o estudo da relação da concentração de pólen com esses demais fatores. As variáveis explicativas consideradas são: indicador de chuva (1 para chuva com duração de pelo menos 3 horas ou chuva breve, porém intensa, ou 0 caso contrário), temperatura em graus Fahrenheit, previsão da velocidade do vento para o dia seguinte (*mph*) e os dias da estação.

Figura 7: Gráficos da raiz da concentração e pólen por temperatura (a), velocidade do vento (b), dias da estação (c)



Fonte: Elaborada pelo autor (2021).

Analisando a Figura 7 é possível notar que a variável dias da estação aparenta possuir relação não linear com a raiz da concentração de pólen de ambrosia, sugerindo um ajuste não paramétrico para esta variável. Ademais, é possível notar também que a raiz da concentração de pólen tende a aumentar conforme a temperatura aumenta, e nenhuma tendência pôde ser observada entre a velocidade do vento e a raiz da concentração e pólen.

Como as observações de concentração de pólen de ambrosia são valores reais positivos, as distribuições para o ajuste via MLGS que podem ser usadas e foram estudadas nessa monografia serão as distribuições Normal, Gama e Inversa Gaussiana. A Tabela 7 apresenta os valores de α , graus de liberdade e critério de informação bayesiano (BIC) para o ajuste de MLGS considerando o método de seleção de nós de Eilers e Marks (1996) com 23 nós, as distribuições Gama com ligação log, Normal com ligação identidade e log e Inversa Gaussiana com ligação log. Para as distribuições Gama com ligação identidade e inversa, Normal com ligação inversa e Inversa Gaussiana com ligação inversa quadrada e identidade houveram problemas de convergência ou inadequação da função de ligação, por isso não foram consideradas.

Tabela 7: α e seu grau de liberdade ($GL(\alpha)$) e BIC para cada método de estimação aplicados

| | Ligação | Normal | | Gama | Inversa Gaussiana |
|--------------|---------|------------|----------|----------|-------------------|
| | | Identidade | Log | Log | Log |
| α | | 3,000 | 3,000 | 0,500 | 0,400 |
| $GL(\alpha)$ | | 11,655 | 12,052 | 22,255 | 23,440 |
| BIC | | 1548,436 | 1474,087 | 1457,597 | 1707,102 |

Fonte: Elaborada pelo autor (2021).

Com base na Tabela 7, pelo BIC, o melhor modelo para o ajuste será o MLGS com distribuição Gama e ligação log. As estimativas para este modelo estão apresentadas na Tabela 8 e a curva não paramétrica na Figura 8, com curva estimada em azul e bandas de confiança em linhas tracejadas.

Tabela 8: Estimativas calculados com o modelo Gama com ligação log

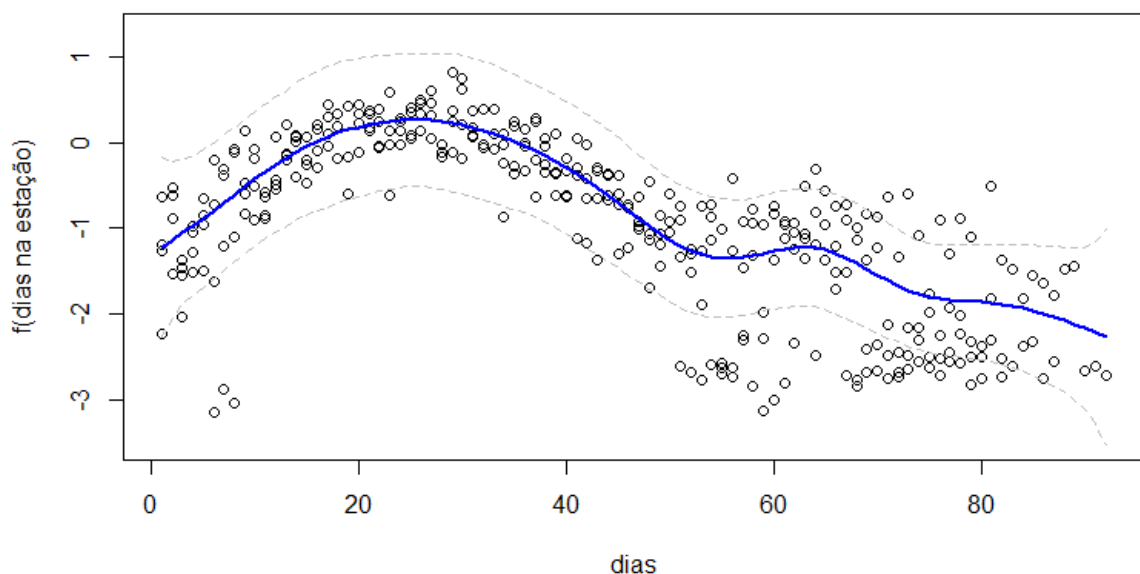
| | Estimativas | Erro padrão |
|-------------------------|-------------|-------------|
| β_1 (chuva) | 0,216 | 0,096 |
| β_2 (temperatura) | 0,022 | 0,004 |
| β_3 (vento) | 0,057 | 0,009 |
| ϕ | 3,844 | – |

Fonte: Elaborada pelo autor (2021).

Portanto o modelo final ajustado foi:

$$\log(\hat{\mu}) = 0,216 * chuva + 0,022 * temperatura + 0,057 * vento + \hat{f}(dias).$$

Figura 8: Curva não paramétrica estimada



Fonte: Elaborada pelo autor (2021).

4 CONCLUSÕES

Neste trabalho foram expandidos os Modelos Lineares Generalizados para a classe semiparamétrica. Foram desenvolvidos detalhadamente os cálculos da função escore e matriz de informação, do método de estimação baseado no método iterativo de Escore de Newton, e do GCV. Foi desenvolvido também, para o *software* estatístico R (R Core Team, 2021), toda a programação de estimação do MLGS considerando as principais distribuições da família exponencial e suas respectivas funções de ligação, além dos métodos de seleção de nós de Green e Silverman (1994) e de Eilers e Marks (1996) e seleção do parâmetro de suavização através da minimização do GCV. O algoritmo foi testado através de estudos de simulação que aferiram a integridade dos cálculos propostos e estão disponíveis na plataforma github pelo *link* <http://www.github.com/ClecioFerreira/SGLM> e *ad request* pelo e-mail: higor.cruz@ice.ufjf.br.

Para trabalhos futuros, uma alternativa é a expansão dos algoritmos de estimação para os Modelos Aditivos Generalizados, onde é considerada mais de uma curva não paramétrica na componente sistemática do MLG, uma vez que todos os cálculos aqui propostos consideraram apenas uma curva não paramétrica. Uma outra possibilidade seria o desenvolvimento de um pacote para o *software* estatístico R (R Core Team, 2021) com os algoritmos de estimação do MLGS, além de gráficos de envelope simulados e medidas de adequação da função de ligação.

5 REFERÊNCIAS

BOWMAN, Adrian W. **An alternative method of cross-validation for the smoothing of density estimates.** *Biometrika* 71, p. 353-360. 1984.

EILERS, Paul HC; MARX, Brian D. **Flexible smoothing with B-splines and penalties.** *Statistical science*, p. 89-102, 1996.

GOLUB, GENE H; HEATH, MICHAEL; WAHBA, GRACE. **Generalized cross-validation as a method for choosing a good ridge parameter.** *Technometrics*, v. 21, n. 2, p. 215-223, 1979.

HARDLE, W. & MARRON, J. S.. **Optimal bandwidth selection in nonparametric regression function estimation.** *Ann. The Annals of Statistics*. 13, p. 1465-1481. 1985.

GREEN, Peter J.; SILVERMAN, Bernard W. **Nonparametric regression and generalized linear models: a roughness penalty approach.** Chapman and Hall, 1994.

McCullagh, P; Nelder, J.A. **Generalized Linear Models.** 2nd Edition, Chapman and Hall, London. 1989.

STONE, M. **Cross-validatory choice and assessment of statistical predictions.** *Journal of the Royal Statistical Society. B* 36, p. 111-147. 1974.

STONE, M. **Cross validation:** a review. *Series Statistics* 9, p. 127-139. 1977,

PAULA, Gilberto A. **Modelos de regressão:** com apoio computacional. São Paulo: IME-USP, 2004.

R Development Core Team (2021). **R:** A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ruppert, D; Wand, M.P; Carrol, R.. **Semiparametric Regression.** Cambridge University Press, New York, 2003.

Stark, P.C., Ryan, L.M., McDonald, J.L. et al. **Using meteorologic data to predict daily ragweed pollen levels.** *Aerobiologia* 13, p 177-184. 1997.