

# CE043 - GAMLSS

## Estimação e Inferência I

Silva, J.P; Taconeli, C.A.

03 de agosto, 2020

# Conteúdo

- 1 Introdução
- 2 Definição dos GAMLSS
- 3 Inferência baseada em verossimilhança
- 4 Propriedades dos estimadores de máxima verossimilhança
- 5 Algoritmos GAMLSS

# Introdução

- Neste módulo vamos tratar da produção de inferências baseadas no ajuste de um modelo da classe GAMLSS;

- Neste módulo vamos tratar da produção de inferências baseadas no ajuste de um modelo da classe GAMLSS;
- Dentre os tópicos a serem abordados, a obtenção de erros padrões, intervalos de confiança, testes de hipóteses e predições;

- Neste módulo vamos tratar da produção de inferências baseadas no ajuste de um modelo da classe GAMLSS;
- Dentre os tópicos a serem abordados, a obtenção de erros padrões, intervalos de confiança, testes de hipóteses e predições;
- As inferências são baseadas, predominantemente, na teoria da verossimilhança.

- Vamos apresentar a formulação geral do modelo GAMLSS;

- Vamos apresentar a formulação geral do modelo GAMLSS;
- Descrever os dois algoritmos para maximizar a log-verossimilhança penalizada com respeito aos parâmetros de efeitos fixos e aos parâmetros de efeitos aleatórios;



- Vamos apresentar a formulação geral do modelo GAMLSS;
- Descrever os dois algoritmos para maximizar a log-verossimilhança penalizada com respeito aos parâmetros de efeitos fixos e aos parâmetros de efeitos aleatórios;
- Explorar função `gamlss` e algumas das funções associadas, e extrair informação de um objeto `gamlss` ajustado.

# Definição dos GAMLSS

# Definição dos GAMLSS

O modelo GAMLSS *semi-paramétrico* aditivo é definido como:

$$\begin{aligned} \mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + \dots + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + \dots + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + \dots + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + \dots + s_{4J_4}(x_{4J_4}) \end{aligned} \tag{1}$$

# Definição dos GAMLSS

O modelo GAMLSS *semi-paramétrico* aditivo é definido como:

$$\begin{aligned} \mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + \dots + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + \dots + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + \dots + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + \dots + s_{4J_4}(x_{4J_4}) \end{aligned} \tag{1}$$

em que

- $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é a distribuição da variável resposta  $\mathbf{Y}$ ,

# Definição dos GAMLSS

O modelo GAMLSS *semi-paramétrico* aditivo é definido como:

$$\begin{aligned} \mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + \dots + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + \dots + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + \dots + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + \dots + s_{4J_4}(x_{4J_4}) \end{aligned} \tag{1}$$

em que

- $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é a distribuição da variável resposta  $\mathbf{Y}$ ,
- $\mathbf{X}_k$  são matrizes de desenho que incorporam os termos aditivos lineares,

# Definição dos GAMLSS

O modelo GAMLSS *semi-paramétrico* aditivo é definido como:

$$\begin{aligned} \mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + \dots + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + \dots + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + \dots + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + \dots + s_{4J_4}(x_{4J_4}) \end{aligned} \tag{1}$$

em que

- $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é a distribuição da variável resposta  $\mathbf{Y}$ ,
- $\mathbf{X}_k$  são matrizes de desenho que incorporam os termos aditivos lineares,
- $\boldsymbol{\beta}_k$  são os parâmetros dos coeficientes lineares, e

# Definição dos GAMLSS

O modelo GAMLSS *semi-paramétrico* aditivo é definido como:

$$\begin{aligned} \mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + \dots + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + \dots + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + \dots + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + \dots + s_{4J_4}(x_{4J_4}) \end{aligned} \quad (1)$$

em que

- $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é a distribuição da variável resposta  $\mathbf{Y}$ ,
- $\mathbf{X}_k$  são matrizes de desenho que incorporam os termos aditivos lineares,
- $\boldsymbol{\beta}_k$  são os parâmetros dos coeficientes lineares, e
- $s_{kj}(x_{kj})$  representam funções suaves para as variáveis explicativas  $x_{kj}$ , para  $k = 1, 2, 3, 4$  e  $j = 1, \dots, J_k$ .

Um caso particular do modelo (1) em que não há termos aditivos em nenhum dos parâmetros é dado por

$$\begin{aligned} \mathbf{Y} &\overset{iid}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 \end{aligned} \tag{2}$$



# Definição dos GAMLSS

Um caso particular do modelo (1) em que não há termos aditivos em nenhum dos parâmetros é dado por

$$\begin{aligned} \mathbf{Y} &\overset{iid}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1 \boldsymbol{\beta}_1 \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2 \boldsymbol{\beta}_2 \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3 \boldsymbol{\beta}_3 \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4 \boldsymbol{\beta}_4 \end{aligned} \tag{2}$$

Este é o modelo GAMLSS *paramétrico*.

# Definição dos GAMLSS

O modelo (1) pode ser generalizado para um GAMLSS *de efeitos aleatórios*:

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\gamma} &\stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_{11}\boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1}\boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_{21}\boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2}\boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_{31}\boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3}\boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + \mathbf{Z}_{41}\boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4}\boldsymbol{\gamma}_{4J_4} \end{aligned} \tag{3}$$

# Definição dos GAMLSS

O modelo (1) pode ser generalizado para um GAMLSS *de efeitos aleatórios*:

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\gamma} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_{11}\boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1}\boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_{21}\boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2}\boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_{31}\boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3}\boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + \mathbf{Z}_{41}\boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4}\boldsymbol{\gamma}_{4J_4} \end{aligned} \tag{3}$$

em que

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T$ , são os parâmetros de efeitos fixos.

# Definição dos GAMLSS

O modelo (1) pode ser generalizado para um GAMLSS *de efeitos aleatórios*:

$$\begin{aligned} \mathbf{Y} | \boldsymbol{\gamma} &\stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_{11} \boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1} \boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_{21} \boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2} \boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3 \boldsymbol{\beta}_3 + \mathbf{Z}_{31} \boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3} \boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4 \boldsymbol{\beta}_4 + \mathbf{Z}_{41} \boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4} \boldsymbol{\gamma}_{4J_4} \end{aligned} \tag{3}$$

em que

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T$ , são os parâmetros de efeitos fixos.
- $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}^T, \dots, \boldsymbol{\gamma}_{1J_1}^T, \boldsymbol{\gamma}_{21}^T, \dots, \boldsymbol{\gamma}_{4J_4}^T)^T$ , são os parâmetros de efeitos aleatórios.

# Definição dos GAMLSS

O modelo (1) pode ser generalizado para um GAMLSS *de efeitos aleatórios*:

$$\begin{aligned} \mathbf{Y} | \boldsymbol{\gamma} &\stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_{11} \boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1} \boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_{21} \boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2} \boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3 \boldsymbol{\beta}_3 + \mathbf{Z}_{31} \boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3} \boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4 \boldsymbol{\beta}_4 + \mathbf{Z}_{41} \boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4} \boldsymbol{\gamma}_{4J_4} \end{aligned} \tag{3}$$

em que

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T$ , são os parâmetros de efeitos fixos.
- $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}^T, \dots, \boldsymbol{\gamma}_{1J_1}^T, \boldsymbol{\gamma}_{21}^T, \dots, \boldsymbol{\gamma}_{4J_4}^T)^T$ , são os parâmetros de efeitos aleatórios.
- Assume-se  $\boldsymbol{\gamma}_{kj} \sim N(\mathbf{0}, [\mathbf{G}(\boldsymbol{\lambda}_{kj})]^{-1})$ , em que  $\boldsymbol{\lambda}_{kj}$  regula a quantidade de suavização do ajuste.

# Definição dos GAMLSS

A função de log-verossimilhança para o modelo (2), sob a suposição de que as observações são independentes, é dada por

$$l = \sum_{i=1}^n \log f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i). \quad (4)$$

# Definição dos GAMLSS

A função de log-verossimilhança para o modelo (2), sob a suposição de que as observações são independentes, é dada por

$$l = \sum_{i=1}^n \log f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i). \quad (4)$$

A função de log-verossimilhança penalizada para o modelo (3) é dado por

$$l_p = l - \frac{1}{2} \sum_{k=1}^4 \sum_{j=1}^{J_k} \gamma_{kj}^T \mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj}) \gamma_{kj}. \quad (5)$$

# Definição dos GAMLSS

A função de log-verossimilhança para o modelo (2), sob a suposição de que as observações são independentes, é dada por

$$l = \sum_{i=1}^n \log f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i). \quad (4)$$

A função de log-verossimilhança penalizada para o modelo (3) é dado por

$$l_p = l - \frac{1}{2} \sum_{k=1}^4 \sum_{j=1}^{J_k} \gamma_{kj}^T \mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj}) \gamma_{kj}. \quad (5)$$

Na sequência, discutiremos a inferência geral com base na função de verossimilhança, propriedades dos estimadores, e discutiremos brevemente os algoritmos de estimação dos parâmetros

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})^T.$$



# Inferência baseada em verossimilhança

# Função de verossimilhança

- A função de verossimilhança, denotada por  $L(\boldsymbol{\theta})$ , é a probabilidade de observar uma amostra ( $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ ) tomada como função dos parâmetros ( $\boldsymbol{\theta}$ );

# Função de verossimilhança

- A função de verossimilhança, denotada por  $L(\boldsymbol{\theta})$ , é a probabilidade de observar uma amostra ( $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ ) tomada como função dos parâmetros ( $\boldsymbol{\theta}$ );
- Considere  $y_1, y_2, \dots, y_n$  variáveis aleatórias discretas, com função de probabilidade  $f(y|\boldsymbol{\theta})$ . Assumindo que as observações são independentes, temos:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta})$$

# Função de verossimilhança

- A função de verossimilhança, denotada por  $L(\boldsymbol{\theta})$ , é a probabilidade de observar uma amostra ( $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ ) tomada como função dos parâmetros ( $\boldsymbol{\theta}$ );
- Considere  $y_1, y_2, \dots, y_n$  variáveis aleatórias discretas, com função de probabilidade  $f(y|\boldsymbol{\theta})$ . Assumindo que as observações são independentes, temos:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta})$$

- Na prática é comum considerar a função de log-verossimilhança, dada por:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_i|\boldsymbol{\theta})$$

# Função de verossimilhança

- No caso de variáveis aleatórias contínuas temos  $P(Y = y) = 0$  para qualquer  $y$ ;

# Função de verossimilhança

- No caso de variáveis aleatórias contínuas temos  $P(Y = y) = 0$  para qualquer  $y$ ;
- Uma versão ligeiramente diferente da função de verossimilhança para variáveis contínuas pode ser definida considerando a probabilidade de uma observação na vizinhança de  $y_i$ , do tipo  $y_i \pm \Delta_i$ ;

# Função de verossimilhança

- No caso de variáveis aleatórias contínuas temos  $P(Y = y) = 0$  para qualquer  $y$ ;
- Uma versão ligeiramente diferente da função de verossimilhança para variáveis contínuas pode ser definida considerando a probabilidade de uma observação na vizinhança de  $y_i$ , do tipo  $y_i \pm \Delta_i$ ;
- A função de verossimilhança, neste caso, fica definida por:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n P(y_i - \Delta_i < Y < y_i + \Delta_i | \boldsymbol{\theta}) \\ &= \prod_{i=1}^n [F(y_i + \Delta_i | \boldsymbol{\theta}) - F(y_i - \Delta_i | \boldsymbol{\theta})] \end{aligned}$$

# Função de verossimilhança

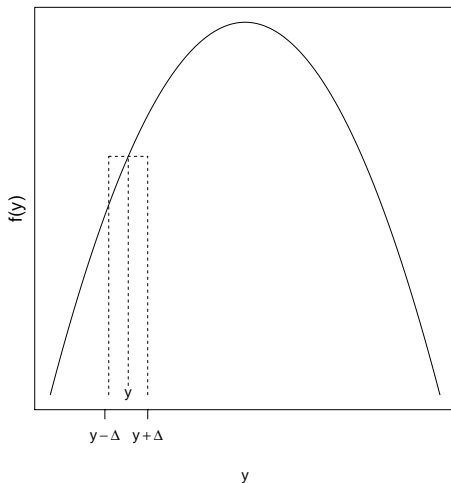


Figura 1: Ilustração - aproximação da verossimilhança para variáveis aleatórias contínuas



- Assumindo que os  $\Delta_i$ 's são suficientemente pequenos, a função de verossimilhança fica definida por:

$$L(\boldsymbol{\theta}) \approx 2 \prod_{i=1}^n f(y_i|\boldsymbol{\theta}) \Delta_i = 2 \left( \prod_{i=1}^n f(y_i|\boldsymbol{\theta}) \right) \left( \prod_{i=1}^n \Delta_i \right)$$

# Função de verossimilhança

- Assumindo que os  $\Delta_i$ 's são suficientemente pequenos, a função de verossimilhança fica definida por:

$$L(\boldsymbol{\theta}) \approx 2 \prod_{i=1}^n f(y_i|\boldsymbol{\theta}) \Delta_i = 2 \left( \prod_{i=1}^n f(y_i|\boldsymbol{\theta}) \right) \left( \prod_{i=1}^n \Delta_i \right)$$

- Já a log-verossimilhança é dada por:

$$l(\boldsymbol{\theta}) \approx \sum_{i=1}^n \log f(y_i|\boldsymbol{\theta}) + \sum_{i=1}^n \log \Delta_i + \log 2$$

# Estimação por máxima verossimilhança

- O método de máxima verossimilhança consiste em encontrar  $\theta$  que maximiza a função de verossimilhança (ou a log-verossimilhança);

# Estimação por máxima verossimilhança

- O método de máxima verossimilhança consiste em encontrar  $\boldsymbol{\theta}$  que maximiza a função de verossimilhança (ou a log-verossimilhança);
- No caso de variáveis contínuas, apenas o primeiro termo, que envolve  $f(y_i|\boldsymbol{\theta})$ , depende de  $\boldsymbol{\theta}$ , de modo que os demais termos podem ser ignorados.

# Estimação por máxima verossimilhança

- O método de máxima verossimilhança consiste em encontrar  $\theta$  que maximiza a função de verossimilhança (ou a log-verossimilhança);
- No caso de variáveis contínuas, apenas o primeiro termo, que envolve  $f(y_i|\theta)$ , depende de  $\theta$ , de modo que os demais termos podem ser ignorados.
- Assim, de maneira geral, os estimadores de máxima verossimilhança são aqueles que maximizam

$$l(\theta) = \sum_{i=1}^n \log f(y_i|\theta),$$

em que  $f(y_i|\theta)$  é uma função de probabilidades, no caso discreto, ou uma função densidade de probabilidades, no caso contínuo.

# Estimação por máxima verossimilhança

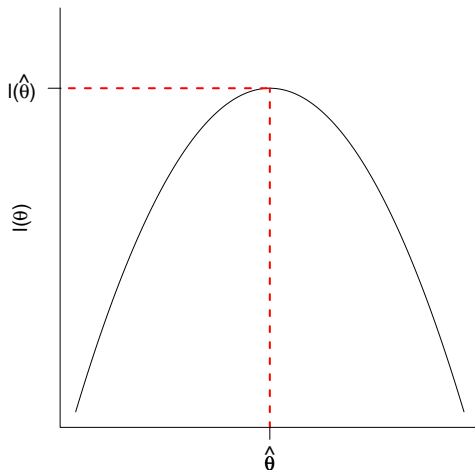


Figura 2: Estimação por máxima verossimilhança

# Propriedades dos estimadores de máxima verossimilhança

# Propriedades dos estimadores de máxima verossimilhança

- **Invariância:** Se  $\hat{\theta}$  é o EMV de  $\theta$ , e  $\phi = g(\theta)$  é uma transformação um a um de  $\theta$ , então  $\hat{\phi} = g(\hat{\theta})$  é o EMV de  $\phi$ ;



# Propriedades dos estimadores de máxima verossimilhança

- **Invariância:** Se  $\hat{\theta}$  é o EMV de  $\theta$ , e  $\phi = g(\theta)$  é uma transformação um a um de  $\theta$ , então  $\hat{\phi} = g(\hat{\theta})$  é o EMV de  $\phi$ ;
- **Consistência:** Sob condições de regularidade,  $\hat{\theta}$  converge em probabilidade para o real parâmetro  $\theta$  quando  $n \rightarrow \infty$ :

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0.$$

# Propriedades dos estimadores de máxima verossimilhança

- **Eficiência assintótica:** Os EMVs são eficientes (apresentam menor erro médio quadrático) em uma grande classe de estimadores quando  $n \rightarrow \infty$ ;

# Propriedades dos estimadores de máxima verossimilhança

- **Eficiência assintótica:** Os EMVs são eficientes (apresentam menor erro médio quadrático) em uma grande classe de estimadores quando  $n \rightarrow \infty$ ;
- **Normalidade assintótica:** Os EMVs apresentam distribuição normal quando  $n \rightarrow \infty$ .

# Propriedades assintóticas dos estimadores

- A inferência baseada na verossimilhança em GAMLSS resulta das propriedades gerais dos EMVs.

# Propriedades assintóticas dos estimadores

- A inferência baseada na verossimilhança em GAMLSS resulta das propriedades gerais dos EMVs.
- Seja  $\boldsymbol{\theta}$  o vetor de parâmetros do modelo, com os coeficientes de regressão ( $\beta$ 's) para  $\mu$ ,  $\sigma$ ,  $\nu$  e  $\tau$ , isto é,  $\boldsymbol{\theta} = (\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\nu, \boldsymbol{\beta}_\tau)$ .

# Propriedades assintóticas dos estimadores

- A inferência baseada na verossimilhança em GAMLSS resulta das propriedades gerais dos EMVs.
- Seja  $\boldsymbol{\theta}$  o vetor de parâmetros do modelo, com os coeficientes de regressão ( $\beta$ 's) para  $\mu$ ,  $\sigma$ ,  $\nu$  e  $\tau$ , isto é,  $\boldsymbol{\theta} = (\beta_\mu, \beta_\sigma, \beta_\nu, \beta_\tau)$ .
- A distribuição assintótica de  $\hat{\boldsymbol{\theta}}$ , o EMV de  $\boldsymbol{\theta}$ , é dada por:

$$\hat{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, I_F(\boldsymbol{\theta})^{-1}\right),$$

em que

$$I_F(\boldsymbol{\theta}) = -E \left[ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right]$$

é a matriz informação esperada de Fisher.

# Matriz de variância-covariâncias assintótica e erros padrões

- Nos casos em que a matriz informação esperada não pode ser obtida, ela pode ser substituída pela matriz informação observada, definida por:

$$I(\boldsymbol{\theta}) = - \left[ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right].$$

# Matriz de variância-covariâncias assintótica e erros padrões

- Nos casos em que a matriz informação esperada não pode ser obtida, ela pode ser substituída pela matriz informação observada, definida por:

$$I(\boldsymbol{\theta}) = - \left[ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right].$$

- Em ambos os casos, a estimativa da matriz de variâncias e covariâncias é estimada substituindo-se  $\boldsymbol{\theta}$  por  $\hat{\boldsymbol{\theta}}$  em  $I_F(\boldsymbol{\theta})$  ou  $I(\boldsymbol{\theta})$ .



# Matriz de variância-covariâncias assintótica e erros padrões

- Nos casos em que a matriz informação esperada não pode ser obtida, ela pode ser substituída pela matriz informação observada, definida por:

$$I(\boldsymbol{\theta}) = - \left[ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right].$$

- Em ambos os casos, a estimativa da matriz de variâncias e covariâncias é estimada substituindo-se  $\boldsymbol{\theta}$  por  $\hat{\boldsymbol{\theta}}$  em  $I_F(\boldsymbol{\theta})$  ou  $I(\boldsymbol{\theta})$ .
- Diferentemente do que ocorre para os GLM's, para os GAMLSS's  $\hat{\boldsymbol{\theta}}$  em geral não é um estimador consistente de  $\boldsymbol{\theta}$  se o modelo não for corretamente especificado.

- Uma alternativa robusta à má especificação do modelo, para a estimação da matriz de variâncias e covariâncias, é dada por:

$$\widehat{Var}(\hat{\boldsymbol{\theta}}) = I(\hat{\boldsymbol{\theta}})^{-1} K(\hat{\boldsymbol{\theta}}) I(\hat{\boldsymbol{\theta}})^{-1},$$

em que  $K(\hat{\boldsymbol{\theta}})$  é uma estimativa amostral da matriz de variâncias e covariâncias.

# Propriedades assintóticas dos estimadores

- Os erros padrões dos parâmetros de regressão são usualmente obtidos tomando a raiz quadrada dos elementos da diagonal da matriz de covariâncias dos estimadores (inversa da matriz informação);

# Propriedades assintóticas dos estimadores

- Os erros padrões dos parâmetros de regressão são usualmente obtidos tomando a raiz quadrada dos elementos da diagonal da matriz de covariâncias dos estimadores (inversa da matriz informação);
- Nos casos em que a obtenção da matriz informação é complicada, uma alternativa para obter o erro padrão de um particular  $\hat{\beta}$  é dada por:

$$\text{e.p.}(\hat{\beta}) \approx \frac{|\hat{\beta}|}{\sqrt{\Delta\text{GDEV}}},$$

em que  $\Delta\text{GDEV}$  é a diferença das *deviances* obtidas com a não inclusão e com a inclusão da variável explicativa correspondente no ajuste do modelo.

# Algoritmos GAMLSS

Há dois algoritmos para estimação em um modelo GAMLSS:

Há dois algoritmos para estimação em um modelo GAMLSS:

- O algoritmo RS, que não usa as derivadas cruzadas da função de log-verossimilhança.

Há dois algoritmos para estimação em um modelo GAMLSS:

- O algoritmo RS, que não usa as derivadas cruzadas da função de log-verossimilhança.
- O algoritmo CG, que requer informação sobre as derivadas primeira, segunda e cruzada de  $\mu$ ,  $\sigma$ ,  $\nu$  e  $\tau$ .



Há dois algoritmos para estimação em um modelo GAMLSS:

- O algoritmo RS, que não usa as derivadas cruzadas da função de log-verossimilhança.
- O algoritmo CG, que requer informação sobre as derivadas primeira, segunda e cruzada de  $\mu$ ,  $\sigma$ ,  $\nu$  e  $\tau$ .

A Figura 3 ilustra como os dois algoritmos alcançam as estimativas de máxima verossimilhança.

- Os contornos se referem à *deviance* global (igual a duas vezes menos a log-verossimilhança).
- As duas figuras foram geradas usando uma amostra aleatória da distribuição Weibull, denotada por  $\text{WEI}(\mu, \sigma)$ .

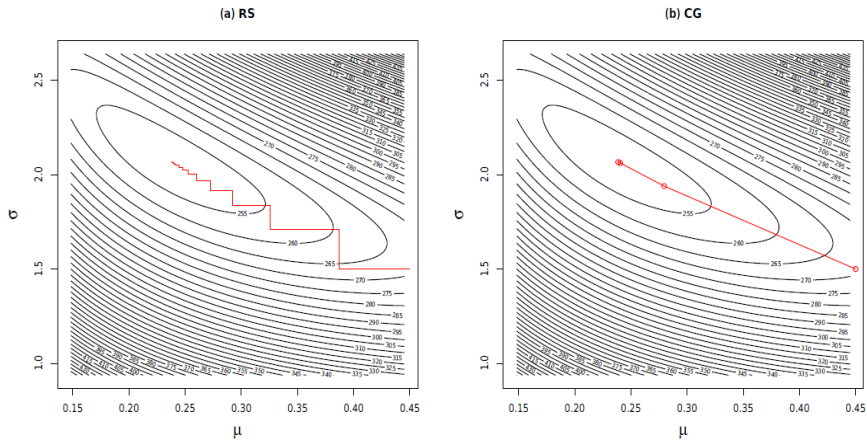


Figura 3: Como os algoritmos (a) RS e (b) CG alcançam as estimativas de máxima verossimilhança

# O algoritmo RS

O algoritmo RS busca a maximização sobre os parâmetros, alternando em cada um até a convergência.

# O algoritmo RS

O algoritmo RS busca a maximização sobre os parâmetros, alternando em cada um até a convergência.

A *deviance* global é minimizada (verossimilhança maximizada) para cada  $\mu$  e  $\sigma$ , alternando até a convergência.

# O algoritmo RS

O algoritmo RS busca a maximização sobre os parâmetros, alternando em cada um até a convergência.

A *deviance* global é minimizada (verossimilhança maximizada) para cada  $\mu$  e  $\sigma$ , alternando até a convergência.

O algoritmo RS é geralmente mais estável e mais rápido, assim é usado por padrão.

# O algoritmo RS

O algoritmo RS busca a maximização sobre os parâmetros, alternando em cada um até a convergência.

A *deviance* global é minimizada (verossimilhança maximizada) para cada  $\mu$  e  $\sigma$ , alternando até a convergência.

O algoritmo RS é geralmente mais estável e mais rápido, assim é usado por padrão.

Quando os parâmetros ajustados são altamente correlacionados o algoritmo RS pode ser mais lento e convergir antes de alcançar o máximo.

# O algoritmo CG

O algoritmo CG tem a habilidade, dado que usa a informação sobre as derivadas cruzadas, de atualizar conjuntamente  $(\mu, \sigma)$ .

# O algoritmo CG

O algoritmo CG tem a habilidade, dado que usa a informação sobre as derivadas cruzadas, de atualizar conjuntamente  $(\mu, \sigma)$ .

Na prática, contudo, o algoritmo CG não é preferível.



# O algoritmo CG

O algoritmo CG tem a habilidade, dado que usa a informação sobre as derivadas cruzadas, de atualizar conjuntamente  $(\mu, \sigma)$ .

Na prática, contudo, o algoritmo CG não é preferível.

Este é muito instável, principalmente no início das iterações, e diverge facilmente.

# O algoritmo CG

O algoritmo CG tem a habilidade, dado que usa a informação sobre as derivadas cruzadas, de atualizar conjuntamente  $(\mu, \sigma)$ .

Na prática, contudo, o algoritmo CG não é preferível.

Este é muito instável, principalmente no início das iterações, e diverge facilmente.

Há a possibilidade de usar uma opção que combina os dois algoritmos.

O algoritmo usado no `gamlss()` é especificado no argumento `method`, com padrão `method=RS()`.

O algoritmo usado no `gamlss()` é especificado no argumento `method`, com padrão `method=RS()`.

O usuário pode especificar `method=CG()`, ou uma combinação dos algoritmos com `method=mixed()`.

O algoritmo usado no `gamlss()` é especificado no argumento `method`, com padrão `method=RS()`.

O usuário pode especificar `method=CG()`, ou uma combinação dos algoritmos com `method=mixed()`.

Esta opção usa RS para as primeiras iterações, mas depois troca para CG, e é recomendada para parâmetros ajustados altamente correlacionados.

# O Algoritmo RS

Utiliza três componentes aninhados:

# O Algoritmo RS

Utiliza três componentes aninhados:

- *Iteração externa*: após a inicialização de  $\hat{\theta} = (\mu_0, \sigma_0, \nu_0, \tau_0)$ , atualiza um parâmetro por vez usando a

Utiliza três componentes aninhados:

- *Iteração externa*: após a inicialização de  $\hat{\theta} = (\mu_0, \sigma_0, \nu_0, \tau_0)$ , atualiza um parâmetro por vez usando a
- *Iteração interna* (algoritmo *scoring*): cada parâmetro ( $\theta_k$ ) é atualizado fazendo ajustes considerando pesos e uma variável resposta modificada. Para GLM's o processo é conhecido como *mínimos quadrados ponderados iterativamente*. Neste passo é chamado o



Utiliza três componentes aninhados:

- *Iteração externa*: após a inicialização de  $\hat{\theta} = (\mu_0, \sigma_0, \nu_0, \tau_0)$ , atualiza um parâmetro por vez usando a
- *Iteração interna* (algoritmo *scoring*): cada parâmetro ( $\theta_k$ ) é atualizado fazendo ajustes considerando pesos e uma variável resposta modificada. Para GLM's o processo é conhecido como *mínimos quadrados ponderados iterativamente*. Neste passo é chamado o
- *Backfitting modificado*: os parâmetros *beta* e *gama* são atualizados. O ajuste dos termos suaves é realizado por meio de algoritmos de mínimos quadrados.

# O Algoritmo RS

Utiliza três componentes aninhados:

- *Iteração externa*: após a inicialização de  $\hat{\theta} = (\mu_0, \sigma_0, \nu_0, \tau_0)$ , atualiza um parâmetro por vez usando a
- *Iteração interna* (algoritmo *scoring*): cada parâmetro ( $\theta_k$ ) é atualizado fazendo ajustes considerando pesos e uma variável resposta modificada. Para GLM's o processo é conhecido como *mínimos quadrados ponderados iterativamente*. Neste passo é chamado o
- *Backfitting modificado*: os parâmetros *beta* e *gama* são atualizados. O ajuste dos termos suaves é realizado por meio de algoritmos de mínimos quadrados.

A convergência ocorre quando há convergência (com base na *deviance*) dos três algoritmos.

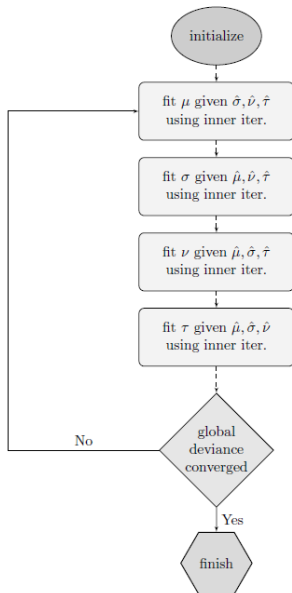


Figura 4: Iteração externa no algoritmo RS.

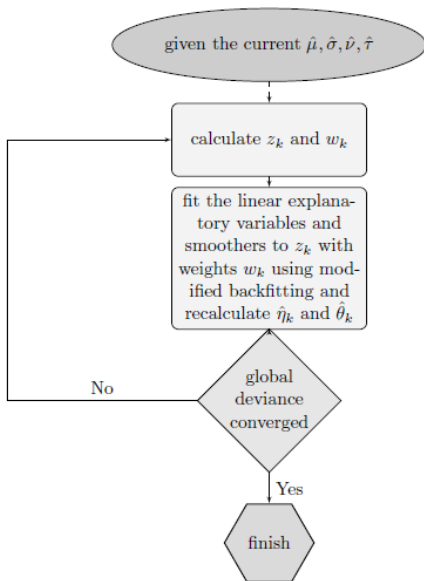


Figura 5: Iteração interna no algoritmo RS.

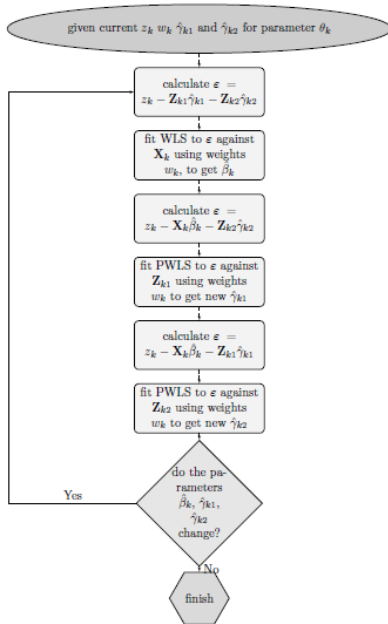


Figura 6: O algoritmo *backfitting* dentro do algoritmo RS

# O Algoritmo CG

O algoritmo CG é um algoritmo *scoring* realizado dentro de uma iteração externa, uma iteração interna e um algoritmo *backfitting* para ajustar cada parâmetro da distribuição.

# O Algoritmo CG

O algoritmo CG é um algoritmo *scoring* realizado dentro de uma iteração externa, uma iteração interna e um algoritmo *backfitting* para ajustar cada parâmetro da distribuição.

Diferente do algoritmo RS, o CG requer as derivadas cruzadas (esperadas ou aproximadas) da log-verossimilhança com relação a cada par de parâmetros da distribuição.

# O Algoritmo CG

O algoritmo CG é um algoritmo *scoring* realizado dentro de uma iteração externa, uma iteração interna e um algoritmo *backfitting* para ajustar cada parâmetro da distribuição.

Diferente do algoritmo RS, o CG requer as derivadas cruzadas (esperadas ou aproximadas) da log-verossimilhança com relação a cada par de parâmetros da distribuição.

Importante citar que o algoritmo RS não é um caso particular do algoritmo CG.



- Modelos paramétricos não lineares podem ser estimados usando a função `nlgamlss()` do pacote `gamlss.nl`.

- Modelos paramétricos não lineares podem ser estimados usando a função `nlgamlss()` do pacote `gamlss.nl`.
- Distribuições adicionais podem ser facilmente adicionadas, já que sua contribuição vem através das derivadas primeira, e derivadas segunda exata ou aproximada (e opcionalmente derivadas cruzadas) e, portanto, é ortogonal ao algoritmo principal.

# Comentários sobre os algoritmos

- Modelos paramétricos não lineares podem ser estimados usando a função `nlglmss()` do pacote `gamlss.nl`.
- Distribuições adicionais podem ser facilmente adicionadas, já que sua contribuição vem através das derivadas primeira, e derivadas segunda exata ou aproximada (e opcionalmente derivadas cruzadas) e, portanto, é ortogonal ao algoritmo principal.
- Valores iniciais são facilmente encontrados, requerendo valores iniciais para  $\theta = (\mu, \sigma, \nu, \tau)$  e não para os parâmetros  $\beta$ .
  - O algoritmo tem se mostrado estável e rápido usando valores iniciais simples (como constantes) para  $\theta$ .
  - Valores *default* podem ser mudados pelo usuário caso necessário.

- Para dado modelo e conjunto de dados específico, a função de log-verossimilhança pode ter múltiplos pontos de máximo. Isso pode ser investigado usando diferentes valores iniciais.

- Para dado modelo e conjunto de dados específico, a função de log-verossimilhança pode ter múltiplos pontos de máximo. Isso pode ser investigado usando diferentes valores iniciais.
- Singularidades na função de verossimilhança podem ocorrer em casos específicos, especialmente quando o tamanho amostral é pequeno.
  - Por exemplo, ocasionalmente o parâmetro de escala  $\sigma$  pode ir para zero. O problema é contornado impondo restrições apropriadas.
  - Assim, ao usar a função de ligação  $\log S$ , uma ligação  $\log$  deslocada de 0.00001, não se permite que ocorram valores menores que 0.00001.