# PGMAT0061/MATG17 - Técnicas Computacionais em Estatística "Simulação Estocástica"

**Professor responsável:** Paulo Henrique Ferreira da Silva

Universidade Federal da Bahia
Instituto de Matemática e Estatística
Departamento de Estatística

21 de Janeiro de 2024

**Otimização Numérica em Estatística:**

Seja $\mathcal{D}$ um domínio fechado e limitado em $\mathbb{R}^S$, e $f(\boldsymbol{x})$ é uma função contínua em $\mathcal{D}$. Queremos encontrar um ponto $\boldsymbol{x}^* \in \mathcal{D}$ tal que:

$$M = f(\boldsymbol{x}^*) = \max_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x})$$

# Métodos de Otimização Numérica

i) Estimação por Máxima Verossimilhança:

$$x_1, \ldots, x_n \ : \text{ amostra aleatória de uma fdp } g(\boldsymbol{x}|\theta),\ \theta \in \Theta \subset \mathbb{R}.$$

$$L(\theta) = \prod_{i=1}^{n} g(x_i|\theta)$$

Queremos encontrar $\widehat{\theta} = \widehat{\theta}_{\mathrm{MV}}$, tal que:

$$L(\widehat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

**OBS.:** Em geral, $\widehat{\theta}$ não tem expressão analítica (ex: Weibull, Beta etc.). Assim, $\widehat{\theta}$ somente pode ser obtido por meio de métodos numéricos.

**PROBLEMA:** Em geral, $L(\theta)$ não é unimodal (máximos locais).

# Métodos de Otimização Numérica

ii) Estatística do Teste da Razão de Verossimilhanças:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \notin \Theta_0 \end{cases}, \qquad \text{com } \Theta_0 \subset \Theta \subset \mathbb{R}$$

<u>TRV:</u> $\lambda = -2 \log \left( \dfrac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \right) \quad \left( \overset{a}{\sim} \chi^2_\nu \right)$

$\nu$: diferença de dimensionalidade de $\Theta_0$ e $\Theta$

# Métodos de Otimização Numérica

iii) Regressão Não Linear:

$$Y_i = h(\boldsymbol{x}_i, \boldsymbol{\theta}) + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

em que $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^S$ é o vetor de parâmetros da regressão (ou coeficientes de regressão) e $h(\cdot)$ é alguma função.

$$\{(Y_i, \boldsymbol{x}_i) \ : \ i = 1, 2, \ldots, n\}$$

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^{n} (Y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta}))^2$$

O estimador de mínimos quadrados $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\text{MQ}}$ de $\boldsymbol{\theta}$ satisfaz:

$$Q(\widehat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta})$$

**OBS:** $\widehat{\boldsymbol{\theta}}$ tem expressão analítica se $h(\cdot)$ é uma função linear de $\boldsymbol{\theta}$. Caso contrário, precisamos usar um método numérico em otimização não linear.

# Métodos de Otimização Numérica

- Métodos de Newton $\Longrightarrow$ usam as derivadas de $1^a$ e $2^a$ ordens da função
- Métodos de Quasi-Newton $\Longrightarrow$ a matriz hessiana de derivadas segundas não precisa ser avaliada diretamente

  Proposto simultaneamente (e de forma independente!) por Broyden, Fletcher, Goldfarb e Shanno em 1970, o método BFGS é um dos mais populares desta classe.

# Método BFGS

Algoritmo:

1. Estabelecer um vetor de valores iniciais $\boldsymbol{x}_0$ e uma matriz hessiana aproximada $\mathrm{B}_0$;
2. Obter uma direção de busca $\boldsymbol{p}_k$ por resolver:
$$\mathrm{B}_k \boldsymbol{p}_k = -\nabla f(\boldsymbol{x}_k)$$
3. Faça uma busca em linha para encontrar um tamanho de passo de descida aceitável $\alpha_k$ na direção encontrada no 2º passo. Então, atualize $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$;
4. Estabeleça $\boldsymbol{s}_k = \alpha_k \boldsymbol{p}_k$ e $\boldsymbol{y}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$;
5. Calcule $\mathrm{B}_{k+1} = \mathrm{B}_k + \dfrac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\boldsymbol{y}_k^\top \boldsymbol{s}_k} - \dfrac{\mathrm{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \mathrm{B}_k^\top}{\boldsymbol{s}_k^\top \mathrm{B}_k \boldsymbol{s}_k}$;
6. Tomar $k = k + 1$;
7. Repetir (2)-(6) enquanto $||\nabla f(\boldsymbol{x}_k)|| > \epsilon$.

# Método BFGS

**OBS.:**

i) $f(\boldsymbol{x})$ é a função objetivo a ser minimizada;

ii) Na prática, $\mathrm{B}_0 = \mathrm{I}$;

iii) No 2º passo do algoritmo, toma-se a inversa da matriz $\mathrm{B}_k$, que pode ser eficientemente obtida ao aplicar a fórmula de Sherman-Morrison no 5º passo do algoritmo, obtendo:

$$\mathrm{B}_{k+1}^{-1} = \left(\mathrm{I} - \frac{\boldsymbol{s}_k \boldsymbol{y}_k^\top}{\boldsymbol{y}_k^\top \boldsymbol{s}_k}\right) \mathrm{B}_k^{-1} \left(\mathrm{I} - \frac{\boldsymbol{y}_k \boldsymbol{s}_k^\top}{\boldsymbol{y}_k^\top \boldsymbol{s}_k}\right) + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^\top}{\boldsymbol{y}_k^\top \boldsymbol{s}_k}.$$

v) No 3º passo, $\alpha_k$ deve satisfazer as condições padrões de Wolfe (1969):

C1) $f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \leq f(\boldsymbol{x}_k) + c_1 \alpha \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k), \quad c_1 \in (0,1);$

C2) $\boldsymbol{p}_k \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \geq c_2 \boldsymbol{p}_k^\top \nabla f(\boldsymbol{x}_k), \quad c_2 \in (0,1) \quad \text{e} \quad 0 < c_1 < c_2 < 1.$

v) Em problemas de estimação na Estatística (tais como EMV e inferência Bayesiana), intervalos de confiança ou de credibilidade podem ser obtidos usando a inversa da matriz hessiana final **(cuidado!!!)**

vi) No R: função optim($\cdot$, method="BFGS")
$\Longrightarrow$ usar control=list(fnscale=-1) p/ torná-lo um problema de maximização!

# Método BFGS

**Exercícios:**

1. $f(x_1, x_2) = x_1^2 + x_2^2 - 8x_1 + 5x_2 + 22{,}25$
2. $f(x_1, x_2) = 10x_1^2 + x_2^2 - 8x_1 + 5x_2 + 22{,}25$
3. $f(x_1, x_2) = 9x_1^2 - 3x_1x_2 + 1{,}25x_2^2 - 24x_1 + 9x_2 + 22{,}25$
4. $f(x_1, x_2) = 0{,}5x_1^2 + 0{,}25x_2^2 + 20\sin(0{,}1x_1x_2) + 0{,}25$

Se possível, representar graficamente a função e a solução encontrada pelo algoritmo.