

Estatística Computacional

Métodos de otimização - Parte 1



Prof. Paulo Cerqueira Jr

Faculdade de Estatística - FAEST

Programa de Pós-graduação em Matemática e Estatística - PPGME

<https://github.com/paulocerqueirajr> 

Introdução

Introdução

- Nessa unidade estaremos interessados no seguinte:

Problema 1: Seja f . Encontre um ponto $\theta \in \Theta$ que minimiza a função f .

- É importante observar que o problema de encontrar um ponto $\theta \in \Theta$ que maximiza uma função $g: \Theta \rightarrow \mathbb{R}$, recai no problema anterior, basta ver que maximizar g é o mesmo que minimizar $f = -g$.
- Problemas de otimização (ou seja, de minimização ou maximização) ocorrem com frequência em diversas áreas das Ciências Exatas, em particular, na Estatística.

Método de Newton-Raphson

Caso unidimensional - Descrição do Método.

- O método de Newton-Raphson é um algoritmo apropriado para encontrar raízes (ou zeros) de funções.
- Formalmente, estamos interessados em encontrar um ponto $\hat{\theta}$ no domínio de uma função $h: \Theta \rightarrow \mathbb{R}$ tal que $h(\hat{\theta}) = 0$.
- Inicialmente vamos considerar o caso onde h é uma função de uma única variável.

Caso unidimensional - Descrição do Método.

- Nessa situação, o método pode ser descrito nos seguintes passos:

1. Fixe um número real $\epsilon > 0$;
2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - \frac{h(\theta_k)}{h'(\theta_k)}.$$

4. Pare o processo iterativo se $|\theta_{k+1} - \theta_k| < \epsilon$. Caso contrário, volte para o passo anterior.

Caso unidimensional - Descrição do Método

- O método de Newton-Raphson possui uma interpretação geométrica simples.
- Basta ver que para todo $k > 0$:

$$h'(\theta_k) = \frac{h(\theta) - 0}{\theta_k - \theta_{k+1}} \Rightarrow \theta_{k+1} = \theta_k - \frac{h(\theta_k)}{h'(\theta_k)}$$

dado alguma aproximação inicial criteriosa θ_0 .

- É possível provar que a sequência $(\theta_k)_{k \geq 0}$ converge para $\hat{\theta}$ quando $k \rightarrow \infty$, se θ_0 é escolhido próximo de $\hat{\theta}$.

Exemplo

A função $h(\theta) = 2\theta - \cos(\theta)$ possui uma raiz real $\hat{\theta}$ isolada no intervalo $[0, \pi/4]$. Encontre um valor aproximado de $\hat{\theta}$ usando o método de Newton-Raphson.

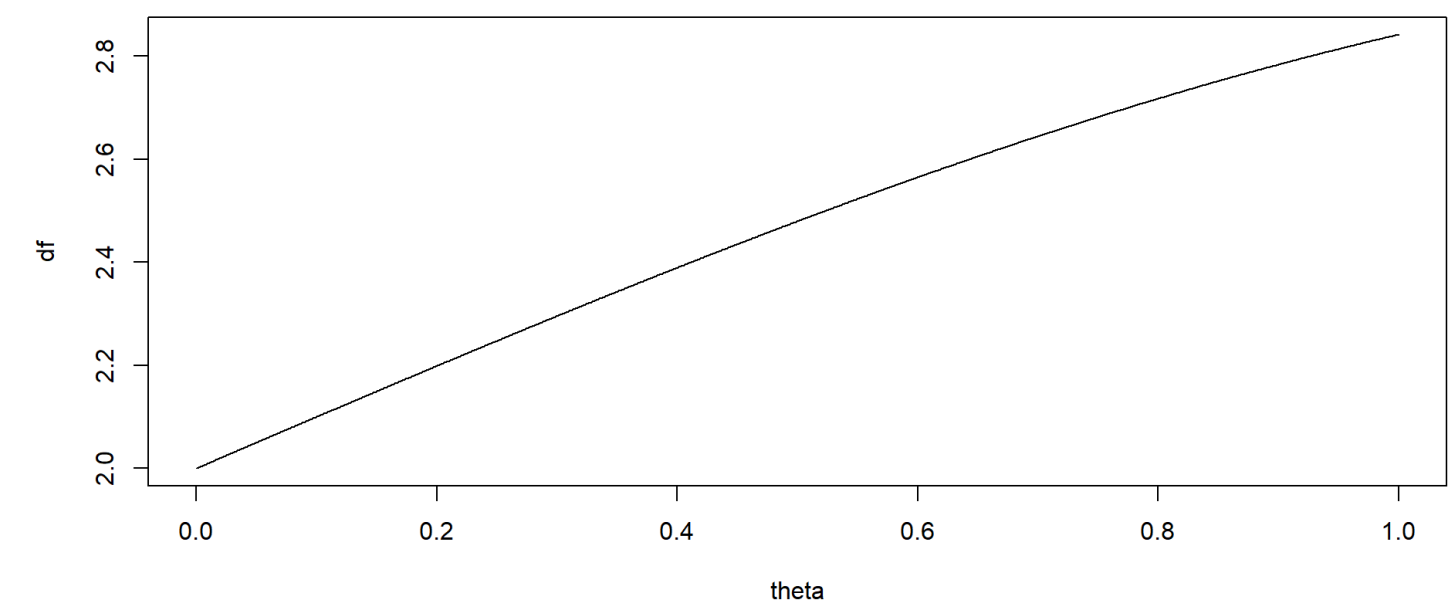
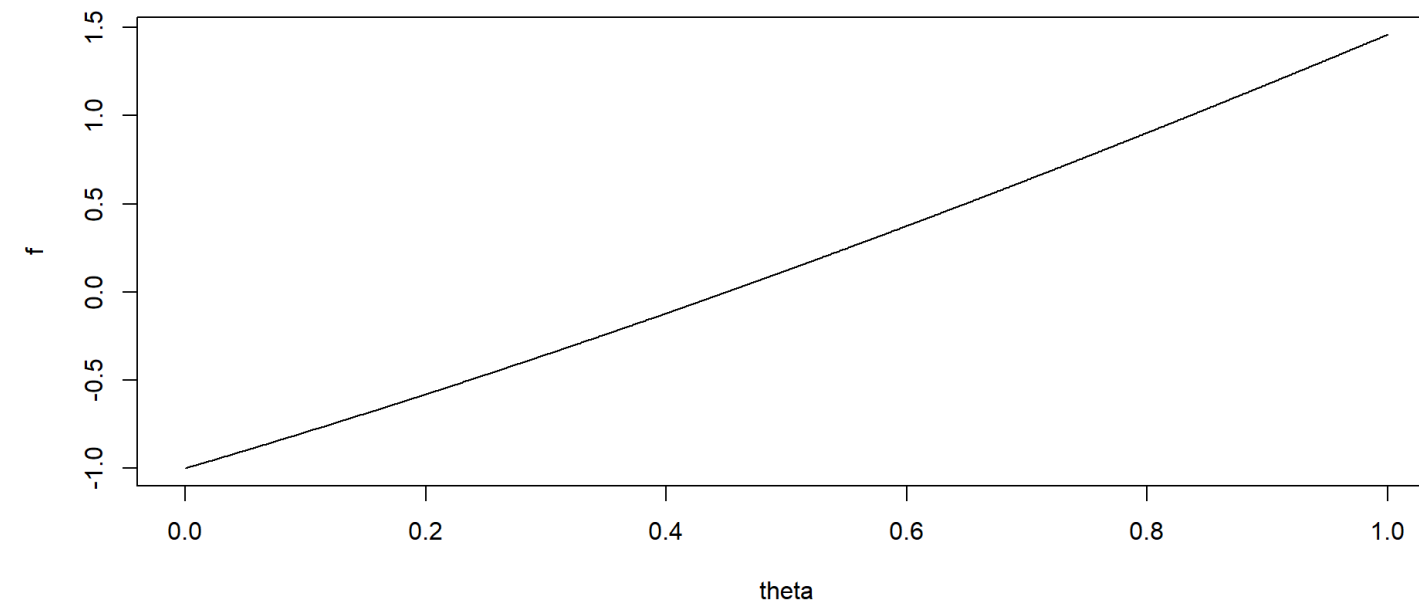
Solução: Primeiramente temos sabemos que:

$$h'(\theta) = 2 + \sin(\theta).$$

- Fixe $\epsilon = 0,0001$ e $\theta_0 = \pi/8$.
- Agora itere e para $k \geq 0$,

$$\theta_{k+1} = \theta_k - \frac{2\theta_k - \cos(\theta_k)}{2 + \sin(\theta_k)}$$

Exemplo



Exemplo:

- Código em R:

```
1 theta.0 <- pi/8
2 theta.0
```

```
[1] 0.3926991
```

```
1 precisao <- 0.0001
2 dif <- 1
3 while(dif > precisao){
4   razao <- (2*theta.0 - cos(theta.0) / sin(theta.0))
5   theta.1 <- theta.0 - razao
6   dif <- abs(theta.1 - theta.0)
7   theta.0 <- theta.1
8   cat("Valor de theta=", theta.0, "\n")
9 }
10 }
```

```
Valor de theta= 0.450819
```

```
Valor de theta= 0.4501837
```

```
Valor de theta= 0.4501836
```

```
1 raiz <- theta.0
2 raiz
```

```
[1] 0.4501836
```

```
1 h <- 2*raiz - cos(raiz)
2 h
```

```
[1] 2.553513e-15
```

Newton-Raphson para Otimização

Newton-Raphson para Otimização

- Considere o problema 1 para o caso em que f é uma função de uma única variável.
- O método de Newton-Raphson é apropriado para resolver numericamente este problema de otimização, basta encontrar as raízes de $h = f'$.
- Neste caso o mínimo θ pode ser encontrado seguindo os seguintes passos:
 1. Fixe um número real $\epsilon > 0$;
 2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
 3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - \frac{f'(\theta_k)}{f''(\theta_k)}.$$

4. Pare o processo iterativo se $|\theta_{k+1} - \theta_k| < \epsilon$. Caso contrário, volte para o passo anterior.

Exemplo

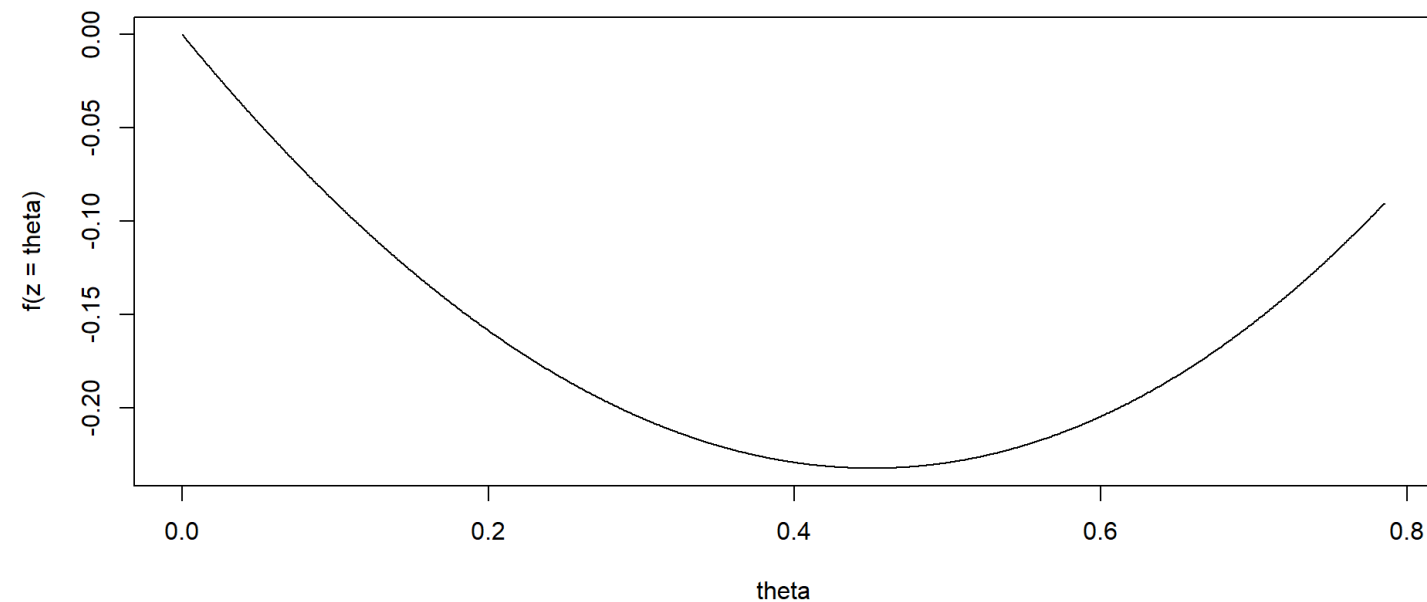
Utilize o método de Newton-Raphson para encontrar o mínimo da função $f(\theta) = \theta^2 - \sin(\theta)$.

Solução:

- Fixe $\epsilon = 0,0001$ e $\theta_0 = \pi/8$, itere e para $k \geq 0$

$$\theta_{k+1} = \theta_k - \frac{2\theta_k - \cos(\theta_k)}{2 + \sin(\theta_k)}$$

Exemplo:



- O **R** também possui funções prontas para pesquisar, dentro de um intervalo, um ponto de mínimo (ou de máximo) de uma função.
- Veja o código abaixo aplicado para o exemplo em questão:

```
1 optimize(f, c(0,pi/4),
2          tol=0.000001) # Otimização
```

\$minimum
[1] 0.4501836

\$objective
[1] -0.2324656

Newton-Raphson em Estatística

Newton-Raphson em Estatística

- Seja (X_1, \dots, X_n) uma a.a. de tamanho n da distribuição de uma v.a. X com densidade $f(x; \theta)$ onde θ pertence ao espaço paramétrico Θ (por enquanto, considere que Θ é unidimensional).
- A função de verossimilhança de θ ($L: \Theta \rightarrow \mathbb{R}$) associada à a.a. observada $\mathbf{x} = (x_1, \dots, x_n)$ é definida por

$$L(\theta) = L(\theta, \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta)$$

- Seja a função de log verossimilhança dada por:

$$\ell(\theta) = \ln(L(\theta)).$$

e a função escore:

$$U(\theta) = \frac{d \ln L(\theta)}{d\theta} = \frac{d\ell(\theta)}{d\theta} = \ell'.$$

Newton-Raphson em Estatística

- Portanto o estimador de máxima verossimilhança, denotado por $\hat{\theta}$, satisfaz as seguintes equações:

$$U(\hat{\theta}) = 0 \quad \Rightarrow \quad \hat{\theta} = \max_{\theta \in \Theta} \ell(\theta).$$

- Em alguns casos pode ser difícil obter uma solução analítica explícita para as equações.
- Nesses casos, é possível obter uma solução aproximada para $\hat{\theta}$ por meio de métodos numéricos.
- Uma alternativa consiste em utilizar o método de Newton-Raphson para aproximar a raiz da função escore (ou maximizar a logverossimilhança).

Newton-Raphson em Estatística

1. Fixe um número real $\epsilon > 0$;
 2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
- Explicitamente, basta seguir o seguinte algoritmo: 3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - \frac{U(\theta_k)}{U'(\theta_k)} = \theta_k - \frac{\ell'(\theta_k)}{\ell''(\theta_k)}.$$

4. Pare o processo iterativo se $|\theta_{k+1} - \theta_k| < \epsilon$. Caso contrário, volte para o passo anterior.

- A sequência $(\theta_k)_{k \geq 0}$ converge para $\hat{\theta}$ quando $k \rightarrow \infty$, se θ_0 é escolhido próximo de $\hat{\theta}$

(Dica: um gráfico de $U(\theta)$ ou $\ell(\theta)$ pode ajudar nessa escolha inicial).

Método Escore

Método Escore

- Em alguns casos, a substituição de $U'(\theta_k)$ por $E(U'(\theta_k))$, apresenta significativa simplificação no procedimento.
- Esse método é conhecido como método do escore e pode ser descrito assim:
 1. Fixe um número real $\epsilon > 0$;
 2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
 3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - \frac{U(\theta_k)}{E(U'(\theta_k))} = \theta_k - \frac{U(\theta_k)}{I(\theta_k)}.$$

em que $I(\theta_k)$ é a informação de Fisher de θ .

Exemplo

Sejam X_1, \dots, X_n uma a.a. de X , com função densidade dada por

$$f(x \mid \theta) = \frac{1}{2}(1 + \theta x), \quad -1 \leq x \leq 1, \quad -1 \leq \theta \leq 1.$$

Determine o EMV para θ pelo método de Newton-Raphson e Escore.

Sol. Inicialmente temos que a função de verossimilhança é dada por

$$L(x \mid \theta) = \frac{1}{2^n} \prod_{i=1}^n (1 + \theta x_i),$$

de modo que

$$U(\theta) = \ell' = \sum_{i=1}^n \frac{x_i}{1 + \theta x_i}$$

Exemplo

E dessa forma

$$U'(\theta) = \ell'' = - \sum_{i=1}^n \frac{x_i^2}{(1 + \theta x_i)^2}.$$

A informação de Fisher de θ é igual,

$$I(\theta) = \frac{1}{2\theta^3} \left\{ \log \left(\frac{1 + \theta}{1 - \theta} \right) - 2\theta \right\}.$$

Gerou-se $n = 20$ valores, com $\theta = 0.4$ usando a função densidade do exemplo via método da transformação inversa, logo

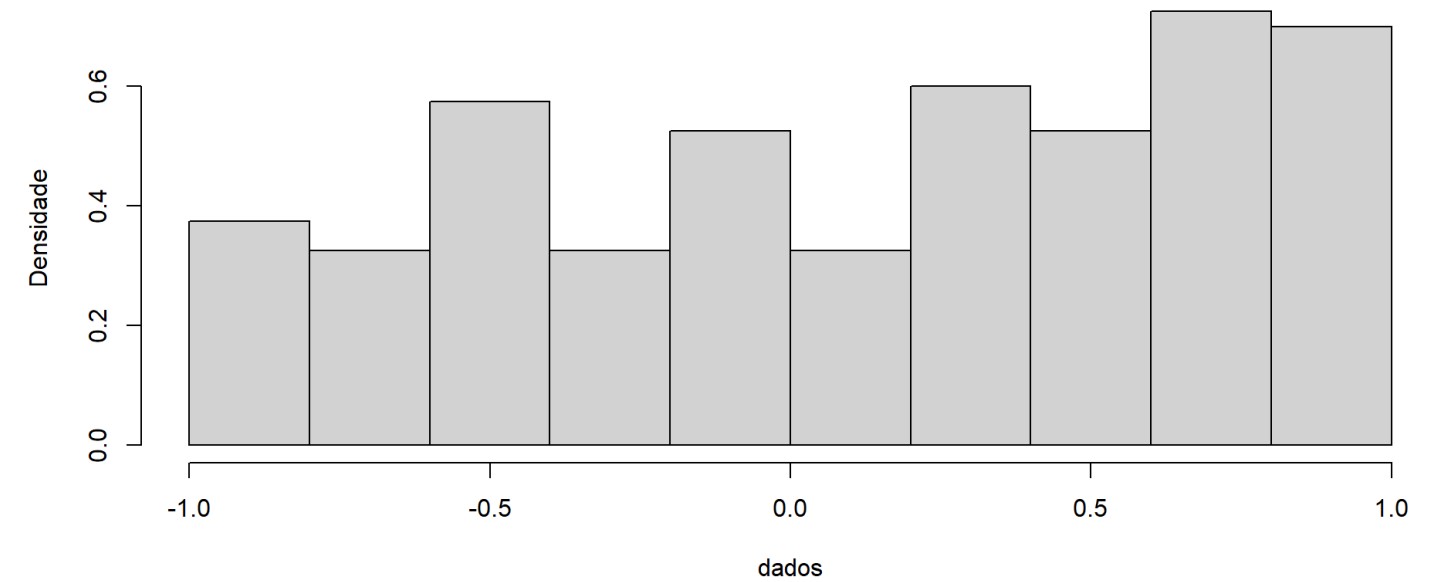
$$x = \frac{-1 + 2\sqrt{1/4 - \theta(1/2 - \theta/4 - u)}}{\theta}.$$

em que $U \sim U(0, 1)$.

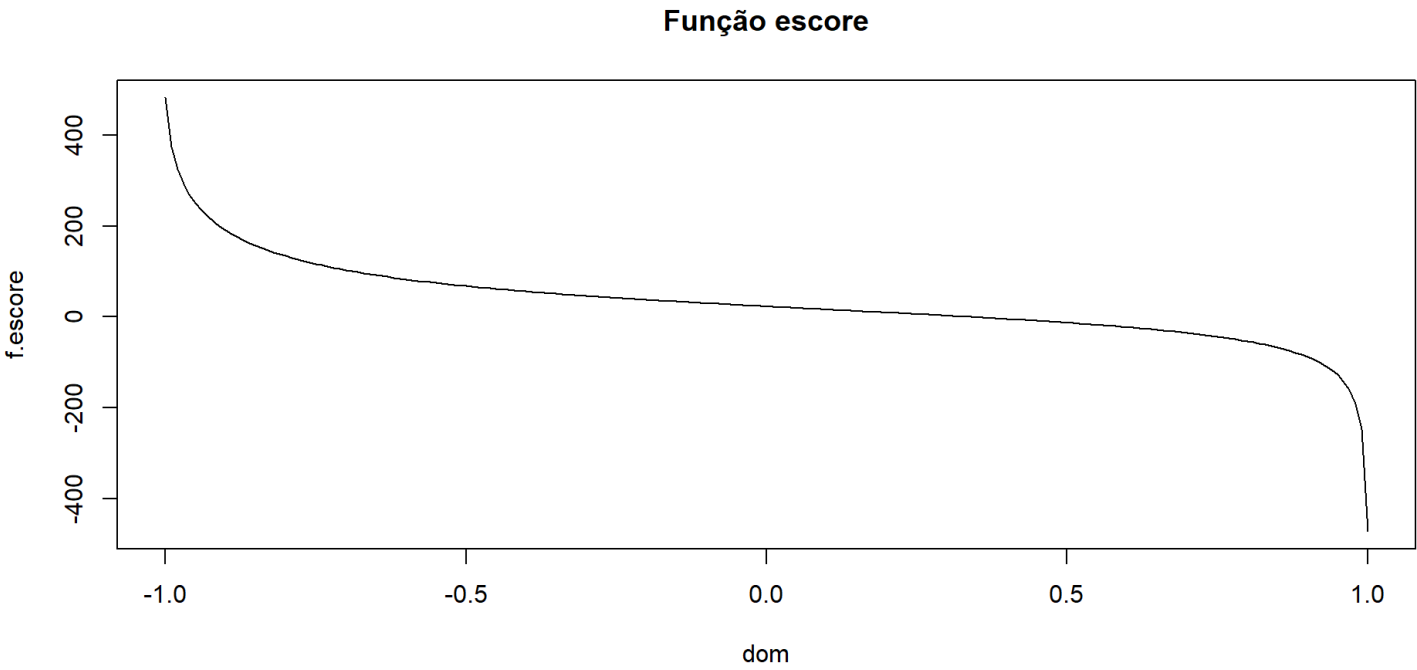
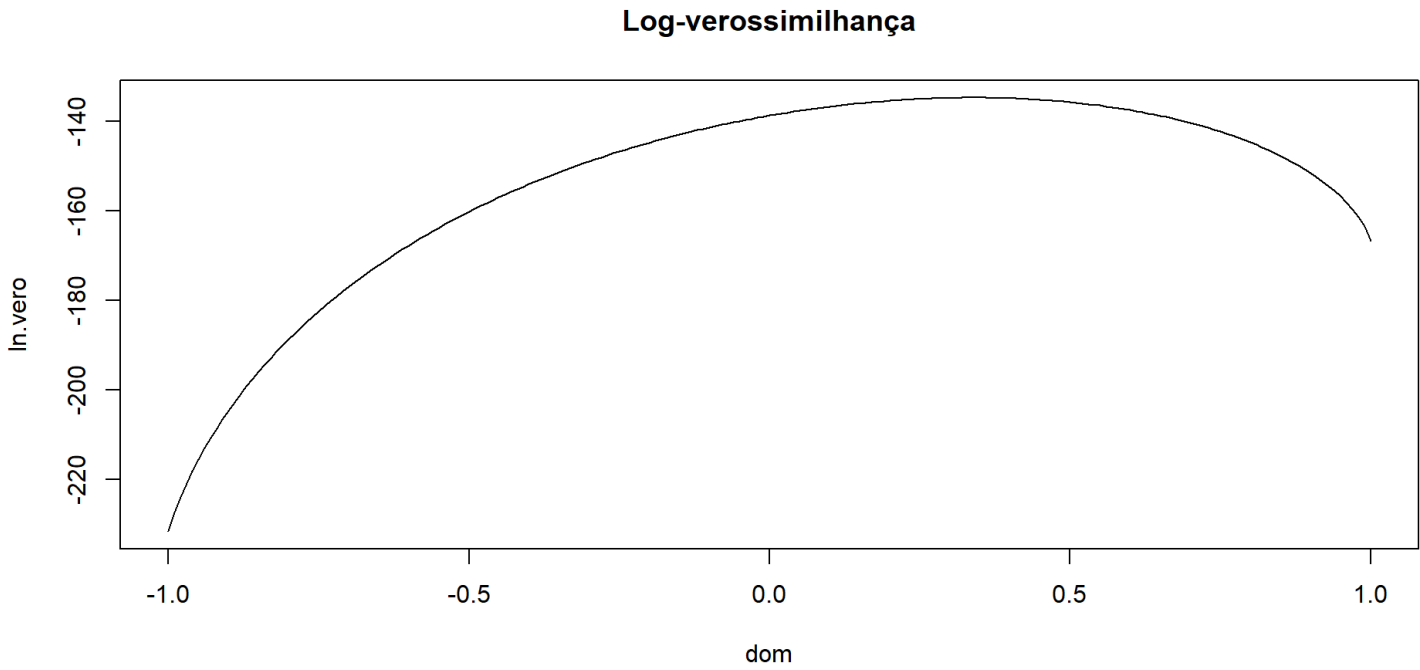
Exemplo

- Código em R:

```
1 set.seed(123456)
2 n <- 200
3 theta <- 0.4
4 u <- runif(n, 0, 1)
5 raiz <- 1 - (theta * (2 - theta)) + (4 * the
6 dados <- (-1 + sqrt(raiz)) / (theta)
```



Exemplo



Exemplo - Comparação dos métodos:

Newton-Raphson:

```
1 theta.zero <- 0.15
2 precisao <- 0.000001
3 dif <- 1
4 while(dif > precisao){
5   num <- S(theta.zero)
6   den <- S.prime(theta.zero)
7   theta.um <- theta.zero - (num/den)
8   dif <- abs(theta.um - theta.zero)
9   theta.zero <- theta.um
10  print(theta.zero)
11 }
```

```
[1] 0.3459363
[1] 0.3398386
[1] 0.3398224
[1] 0.3398224
```

```
1 raiz.NR <- theta.zero
2 raiz.NR # Método NR.
```

```
[1] 0.3398224
```

Escore:

```
1 theta.zero <- 0.15
2 dif <- 1
3 while(dif > precisao){
4   num <- S(theta.zero)
5   a <- 2*theta.zero
6   b <- log((1+theta.zero)/(1-theta.
7   den <- n*(1/(2*theta.zero^3))*b
8   theta.um <- theta.zero + (num/den)
9   dif <- abs(theta.um - theta.zero)
10  theta.zero <- theta.um
11  print(theta.zero)
12 }
```

```
[1] 0.3433802
[1] 0.3397711
[1] 0.3398231
[1] 0.3398223
```

```
1 raiz.E <- theta.zero
2 raiz.E # Método Escore
```

```
[1] 0.3398223
```

Caso Multidimensional

Método de Newton-Raphson

- Agora considere o problema de otimização quando Θ é um espaço multidimensional.
- Antes de apresentar o método de Newton-Raphson nesse caso, vejamos alguns conceitos básicos de Cálculo.

Noções preliminares:

Seja n um inteiro positivo e seja $D \subseteq \mathbb{R}^n$. Considere uma função g que associa a cada $\mathbf{x} = (x_1, \dots, x_n) \in D$ um número real $g(\mathbf{x})$, ou seja, $g: D \rightarrow \mathbb{R}$. O gradiente de g , denotado por

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n} \right).$$

Método de Newton-Raphson

- Seja (X_1, \dots, X_n) uma a.a. de tamanho n da distribuição de uma v.a. X com densidade $f(x; \theta)$ onde $\theta = (\theta_1, \dots, \theta_n)$ pertence ao espaço paramétrico Θ .
- A função de verossimilhança de θ ($L: \Theta \rightarrow \mathbb{R}$) associada à a.a. observada $\mathbf{x} = (x_1, \dots, x_n)$ é definida por

$$L(\theta) = L(\theta, \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta)$$

- Seja a função de log verossimilhança dada por:

$$\ell(\theta) = \ln(L(\theta)).$$

Método de Newton-Raphson

- O i -ésimo elemento do vetor escore, denotado por $U(\theta)$, é dado por

$$U_i(\theta) = \frac{\partial \ell(\theta)}{\partial \theta^{(i)}}.$$

- O (i, j) -elemento da matriz Hessiana, denotada por $H(\theta)$, é dado por

$$H_{ij} = \frac{\partial^2 \ell}{\partial \theta^{(i)} \partial \theta^{(j)}}.$$

Portanto o estimador de máxima verossimilhança, denotado por $\hat{\theta}$, satisfaz as seguintes equações:

$$U(\hat{\theta}) = \nabla \ell(\hat{\theta}) = \mathbf{0} \quad \hat{\theta} = \max_{\theta \in \Theta} \ell(\theta).$$

Newton-Raphson em Estatística

- Em alguns casos pode ser difícil obter uma solução analítica explícita para as equações.
- Nesses casos, é possível obter uma solução aproximada para $\hat{\theta}$ por meio de métodos numéricos.
- Um alternativa consiste em utilizar o método de Newton-Raphson para aproximar a raiz da função escore (ou maximizar a logverossimilhança).

Newton-Raphson em Estatística

1. Fixe um número real $\epsilon > 0$;
2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
- Explicitamente, basta seguir o seguinte algoritmo: 3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - [H(\theta_k)]^{-1} U(\theta_k).$$

4. Pare o processo iterativo se $|\theta_{k+1} - \theta_k| < \epsilon$. Caso contrário, volte para o passo anterior.
- A sequência $(\theta_k)_{k \geq 0}$ converge para $\hat{\theta}$ quando $k \rightarrow \infty$, se θ_0 é escolhido próximo de $\hat{\theta}$

Método Escore

- Por vezes substituir de $H(\theta_k)$ por $E(H(\theta_k))$ pode apresentar significativa simplificação no procedimento.
- Esse método é conhecido como método do escore e pode ser descrito assim:

1. Fixe um número real $\epsilon > 0$;
2. Dê uma aproximação inicial θ_0 para $\hat{\theta}$;
3. Para $k \geq 0$, faça

$$\theta_{k+1} = \theta_k - [E(H(\theta_k))]^{-1} U(\theta_k) = \theta_k - [-I(\theta_k)]^{-1} U(\theta_k),$$

onde $I(\theta_k)$ é a matriz de informação de Fisher de θ .

4. Pare o processo iterativo se $|\theta_{k+1} - \theta_k| < \epsilon$. Caso contrário, volte para o passo anterior.
- A sequência $(\theta_k)_{k \geq 0}$ converge para $\hat{\theta}$ quando $k \rightarrow \infty$, se θ_0 é escolhido próximo de $\hat{\theta}$

Monte Carlo Annealing

Monte Carlo Annealing

- Considere novamente o problema 1 que consiste em encontrar um ponto $\theta \in \Theta$ que minimiza uma função $f: \Theta \rightarrow \mathbb{R}$.
- A ideia fundamental do método de Monte Carlo Annealing (ou Simulated Annealing) para resolver esse problema é emprestada da física.
- Em física da matéria condensada, *annealing* é um processo térmico utilizado para minimizar a energia livre de um sólido.
- Informalmente o processo pode ser descrito em duas etapas:
 - Aumentar a temperatura do sólido até ele derreter;
 - Diminuir lentamente a temperatura até as partículas se organizarem no estado de mínima energia do sólido.

Monte Carlo Annealing

- Esse processo físico pode ser facilmente simulado no computador considerando o algoritmo de Metropolis (1994) cujos passos são:
 1. Fixe uma temperatura inicial T para sólido, um estado inicial θ (onde $\theta \in \Theta$) e a correspondente energia $H(\theta)$ (onde $H: \Theta \rightarrow \mathbb{R}$);
 2. Um estado candidato θ' de energia $H(\theta')$ é gerado aplicando uma pequena perturbação no estado θ . Aceite o estado candidato como o novo estado do sólido conforme a seguinte probabilidade

$$\alpha_T(\theta, \theta') = \begin{cases} 1, & \text{se } H(\theta') - H(\theta) \leq 0 \\ \exp\left(-\frac{H(\theta') - H(\theta)}{T}\right), & \text{se } H(\theta') - H(\theta) > 0 \end{cases}$$

Observe que podemos reescrever $\alpha_T(\theta, \theta')$ da seguinte maneira

Monte Carlo Annealing

3. Repita o passo anterior muitas vezes, considerando sempre a mesma temperatura T ;
4. Diminua a temperatura T e volte para o passo (2).
 - Se o resfriamento é realizado lentamente, o sólido alcança o equilíbrio térmico a cada temperatura.
 - Do ponto de vista da simulação, isso significa gerar muitas transições a uma certa temperatura T .

Monte Carlo Annealing

Para o nosso problema de otimização, faremos a seguinte analogia:

- As soluções do problema de otimização são equivalentes aos estados físicos;
- A função $f: \Theta \rightarrow \mathbb{R}$ é equivalente à função energia H do sólido;
- Um parâmetro de controle $c > 0$ é equivalente à temperatura T .
- Em termos dessa analogia o método de Monte Carlo Annealing pode ser descrito assim:
 1. Escolha $k = 0$, $\theta = \theta_0 \in \Theta$, c_0 e L_0 ;
 2. Faça i de 1 até L_k
 - Gere θ' na vizinhança θ e gere $U \sim U(0, 1)$.
 - Se $f(\theta') \leq f(\theta)$, então $\theta \leftarrow \theta'$
 - Se $f(\theta') > f(\theta)$ e se $U < \exp\left(-\frac{f(\theta') - f(\theta)}{c_k}\right)$ então $\theta \leftarrow \theta'$.
 - Fim do faça.
 3. $k \leftarrow k + 1$
 4. Defina c_k e L_k e volte ao passo 2 até algum critério de parada.

Monte Carlo Annealing

- A sequência $(c_k)_{k \geq 0}$ deve ser escolhida tal que $c_k \rightarrow 0$ lentamente quando $k \rightarrow \infty$.
- Uma boa escolha para c_k , consiste em fazer

$$c_k = \frac{a}{\ln(k+1)}.$$

para alguma constante a apropriada.

- A sequência $(L_k)_{k \geq 0}$ deve ser escolhida tal que para cada valor do parâmetro c_k as soluções (após um transiente inicial) sejam escolhidas de acordo com a seguinte distribuição de probabilidade

$$\pi_{c_k} \propto \exp(-f(\theta)/c_k).$$

Exemplo

Sejam X_1, \dots, X_n uma a.a. de X , com função densidade dada por

$$f(x \mid \theta) = \frac{1}{2}(1 + \theta x), \quad -1 \leq x \leq 1, \quad -1 \leq \theta \leq 1.$$

Determine o EMV para θ pelo método de Newton-Raphson e Escore.