Modelos Lineares Generalizados - Métodos de Estimação

Erica Castilho Rodrigues

17 de Maio de 2017

Introdução

Método de Máxima Verossimilhança

Componentes dos MLG's

- Os MLG's são compostos por duas partes:
  - componente sistemático e componente aleatório.
- **Componente Sistemático**:
  - parte fixa, não aleatória;
  - formada pelo preditor linear e a função de ligação

$$\mu_i = g(\mathbf{x}_i^T \boldsymbol{\beta}) \, .$$

- **Componente Aleatória**:
  - nem toda variação de Y é explicada pelas covariáveis;
  - esse erro não explicado é a Componente Aleatória;
  - em regressão linear são os ϵ_i's;
  - nos MLG's são a distribuição de Y;
  - os Y's estão em torno de μ_i, existe uma variabilidade em torno desse valor.

- Vimos até agora a definição de um Modelo Linear Generalizado.
- Veremos agora como estimar seus parâmetros.
- Como fazemos isso em Regressão Linear?
- Método dos Mínimos Quadrados.
- Pode-se mostrar que é equivalente ao estimador de máxima verossimilhança.
- Para os MLG's também usaremos Método de Máxima Verossimilhança.

Uma outra possibilidade...

- Um método mais simples é o Métodos dos Momentos.
- O que é esse método?
- Igualamos os momentos amostrais aos populacionais.
- Fazemos

$$E(Y) = \overline{y} \quad Var(Y) = S^2 \; .$$

- Vimos que, se a distribuição pertence à família exponencial, sua densidade pode ser escrita na forma

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right] .$$

- No caso canônico essa densidade se reduz a

$$f(y;\theta) = \exp\left[yb(\theta) + c(\theta) + d(\theta)\right] .$$

- Temos então que

$$E(Y) = -\frac{c'(\theta)}{b'(\theta)}$$

$$Var(Y) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$$

- Portanto para encontrarmos o estimador através do Método dos momentos basta fazer

$$-\frac{c'(\theta)}{b'(\theta)} = \overline{y}$$

$$\frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} = S^2$$

- O valor de θ que satisfaz essas equações é o estimador desejado.

- O método dos momentos não gera estimadores muito bons.
- O método mais usado é o Máxima Verossimilhança.
- Podemos usar os valores obtidos como chutes iniciais para outros algoritmos de estimação.

Método de Máxima Verossimilhança

O que é o Método de Máxima Verossimilhança?

Consiste em encontrar o valor do parâmetro que torna mais verossímel a amostra observada.

- Queremos encontrar o valor de θ que maximiza a função de verossimilhança

$$L(\mathbf{y}, \theta) = \prod_{i=1}^{n} f(y_i, \theta) .$$

Exemplo:

- Considere uma amostra aleatória

$$Y_1, Y_2, \ldots, Y_n$$

tal que $Y_i \sim^{iid} Poisson(\theta)$.

- Qual o EMV para θ? $\overline{X}$.
- A função de verossimilhança é dada por

$$f(\mathbf{y}, \theta) = \prod_{i=1}^{n} f(y_i, \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!}$$

- Precisamos maximizar essa função.
- Como isso é feito?
- Derivando e igualando a zero.

Exemplo: (continuação)

- Como podemos facilitar a maximização?
- Tomando o log.

$$\log(f(\mathbf{y},\theta)) = \log\left(\prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!}\right)$$

$$= \sum_i y_i \log(\theta) - \sum_i \theta - \sum_i \log(y_i)$$

$$= \log(\theta) \sum_i y_i - n\theta - \sum_i \log(y_i)\,.$$

- Derivando com relação a θ

$$\frac{d \log(f(\mathbf{y},\theta))}{d\theta} = \frac{\sum_i y_i}{\theta} - n$$

Exemplo: (continuação)

- Igualando a zero

$$\frac{\sum_i y_i}{\theta} - n = 0 \Rightarrow \frac{\sum_i y_i}{\theta} = n$$

$$\hat{\theta} = \frac{\sum_i y_i}{n} = \overline{y}$$

- Então o EMV de θ é a média amostral.

- Em muitos casos não é tão simples obter o EMV.
- Ele pode não ter uma forma fechada.
- Precisamos então de usar métodos numéricos para maximizar a função.
- Uma possibilidade: **Método de Newton Raphson**

Método de Neton Raphson

- É uma poderosa ferramenta para resolver equações numericamente.
- Queremos encontrar o valor de x tal que

$$f(x) = 0 \, .$$

- Se baseia na ideia de:
  - aproximar uma função por uma reta.

- Seja $f(x)$ uma função bem comportada:
  - contínua, possui as primeiras derivadas, etc.
- Vamos denotar por r a raiz da equação

$$f(x) = 0 \,.$$

- Qual nosso objetivo?
- Encontrar o valor de r.
- Começamos com um chute x_0.

- De x_0 vamos para um chute melhor x_1.
- De x_1 produzimos uma nova estimativa x_2
- Esperamos que essa sequencia de números fique cada vez mais próxima de r.
- Vamos melhorando nosso chute até chegar a um valor bem próximo de r.
- Esse é um método iterativo.
- Vejamos agora com detalhes como o algoritmo funciona.

- Iniciamos com o valor x_0.
- Deve ser o máis próximo possível de r.
- Caso contrário, o algoritmo demora demais para encontrar para convergir.
- Para o EMV podemos usar o estimador do Método de Momentos como chute inicial.
- Podemos escrever

$$r = x_0 + h \Rightarrow h = r - x_0$$

  onde h mede o quão longe x_0 está do valor que queremos encontrar.
- Esperamos que o valor de h seja pequeno.

- Como defininos a derivada de uma função em um ponto?

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

- Portanto, se h é pequeno, podemos escrever

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0)}{h} \; .$$

- Vamos isolar o termo $f(x_0 + h)$.
- Estamos interessados nesse termo por $r = x_0 + h$.
- Temos então que

$$hf'(x_0) \approx f(x_0 + h) - f(x_0) \Rightarrow f(x_0 + h) \approx f(x_0) + hf'(x_0)$$

- Isso significa que

$$f(r) = f(x_0 + h) \approx f(x_0) + hf'(x_0) \, .$$

- Mas r é raiz da equação, portanto

$$f(r) = 0 \, .$$

- Substituindo na equação acima

$$f(x_0) + hf'(x_0) \approx 0 \Rightarrow h \approx -\frac{f'(x_0)}{f(x_0)} \, .$$

- Já vimos que
$$r = x_0 + h$$
substituindo $h \approx -\frac{f(x_0)}{f'(x_0)}$ ficamos com
$$r \approx x_0 - \frac{f(x_0)}{f'(x_0)} \, .$$
- Então o nosso próximo chute será
$$x_1 \approx x_0 - \frac{f(x_0)}{f'(x_0)} \, .$$
- E se por sorte nosso chute foi exato?
- Nesse caso $f(x_0) = 0$ o que implica que
$$x_1 = x_0 \, .$$
- Não precisamo continuar chutando.

- Se o nosso chute não acertou de primeira, precisamos continuar.
- O segundo chute será dado por

$$x_1 \approx x_0 - \frac{f(x_0)}{f'(x_0)} \, .$$

- Se $f(x_1) = 0$, o algoritmo para.
- Caso contrário, precisamos de outro chute.
- Qual será o terceiro chute?

$$x_2 \approx x_1 - \frac{f(x_1)}{f'(x_1)} \, .$$

- De maneira geral, no passo n

$$x_n \approx x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \, .$$

- Continuamos até que

$$f(x_n) \approx 0 \, .$$

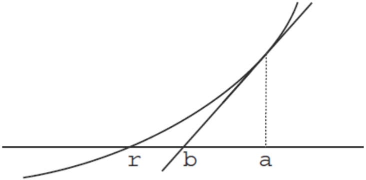
- Provavelmente não chegaremos a um ponto tal que

$$f(x_n) = 0 \, .$$

- Definimos um limiar.
- Por exemplo, paramos se

$$|f(x_n)| < 10^{-5} \, .$$

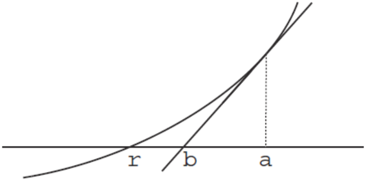
- Vejamos a interpretação geométrica do algoritmo.
- A figura mostra uma função $f(x)$.
- A raz dessa função é o ponto r.
- Nosso primeiro chute é o a.
- Traçamos a reta tangente nesse ponto.



- A reta tangente é dada por

$$y = f(a) + (x - a)f'(a)$$

- O nosso próximo chute é o b.
- Esse é o ponto onde a reta cruza o eixo.
- Fazendo $y = 0$



$$f(a)+(x-a)f'(a) = 0 \rightarrow x = a - \frac{f(a)}{f'(a)}.$$

(mostrar animação no R)

- Vejamos como usar esse algoritmo para encontrar o EMV.
- Queremos encontrar o valor de θ que maximiza $f(\mathbf{y}, \theta)$.
- Como fazemos isso?
- Derivando e igualando a zero.
- Para facilitar, derivamos o log da verossimilhança

$$\log(f(\mathbf{y}, \theta)) = l(\mathbf{y}, \theta)$$

- Queremos descobrir qual valor de θ tal que

$$\frac{dl(\mathbf{y}, \theta)}{d\theta} = 0 \, .$$

- Qual nome dessa função?
- Função Score

$$\frac{dl(\mathbf{y}, \theta)}{d\theta} = U(\theta) \, .$$

- Como usaremos o Newton Raphson?
- Para encontrar a raiz da equação

$$\frac{dl(\mathbf{y}, \theta)}{d\theta} = U(\theta) = 0\ .$$

- Começaremos com um valor inicial θ_0.
- No passo seguinte fazemos

$$\theta_1 = \theta_0 - \frac{U(\theta)}{U'(\theta)}$$

onde

$$U'(\theta) = \frac{dU(\theta)}{d\theta} = \frac{d^2 l(\mathbf{y}, \theta)}{d\theta^2}$$

Algoritmo Newton Raphson

1. Escolha um valor inicial θ_0 (usado método dos momentos, por exemplo).
2. Calcule
$$U(\theta_0) \quad U'(\theta_0)\,.$$
3. Faça
$$\theta_1 = \theta_0 - \frac{U(\theta_0)}{U'(\theta_0)}\,.$$
4. Calcule $U(\theta_1)$.
5. Se $|U(\theta_1)| < 10^{-6}$ o algoritmo para.
6. Caso contrário, volta parao passo 2.

Exemplo:

- Considere uma amostra aleatória

$$Y_1, Y_2, \ldots, Y_n \,.$$

- Suponha que a função densidade dessa variável seja dada por

$$f(y, \theta) = \frac{\theta^y}{y[-\log(1-\theta)]} \,.$$

- A função de verossimilhança fica

$$f(\mathbf{y}, \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i[-\log(1-\theta)]} \,.$$

Exemplo: (continuação)

- Tomando o logaritmo

$$\log(f(\mathbf{y},\theta)) = l(\mathbf{y},\theta) = \sum_{i=1}^{n} \log\left(\frac{\theta^{y_i}}{y_i[-\log(1-\theta)]}\right)$$

$$= \sum_{i=1}^{n} \left(y_i \log(\theta) - \log(y_i) - \log(-\log(1-\theta))\right) .$$

- Derivando com relação a θ, para obter a Função Escore

$$U(\theta) = \frac{dl(\mathbf{y},\theta)}{d\theta} = \sum_{i=1}^{n} \left(\frac{y_i}{\theta} - \frac{1}{-\log(1-\theta)}\left(-\frac{1}{1-\theta}(-1)\right)\right)$$

$$= \frac{dl(\mathbf{y},\theta)}{d\theta} = \sum_{i=1}^{n} \left(\frac{y_i}{\theta} + \frac{1}{\log(1-\theta)}\left(\frac{1}{1-\theta}\right)\right) =$$

$$= \frac{\sum_i y_i}{\theta} + \frac{n}{\log(1-\theta)}\left(\frac{1}{1-\theta}\right)$$

Exemplo: (continuação)

- Derivando novamente com relação a θ

$$\frac{dU(\theta)}{d\theta} = -\frac{\sum_i y_i}{\theta^2} + \frac{d\frac{n}{\log(1-\theta)}}{d\theta}\left(\frac{1}{1-\theta}\right) + \frac{n}{\log(1-\theta)}\frac{d\left(\frac{1}{1-\theta}\right)}{d\theta}$$

$$= -\frac{\sum_i y_i}{\theta^2} + \left(\frac{n(1/(1-\theta))}{log^2(1-\theta)}\right)\left(\frac{1}{1-\theta}\right) + \frac{n}{\log(1-\theta)}\left(-\frac{1}{(1-\theta)^2}(-1)\right)$$

$$= -\frac{\sum_i y_i}{\theta^2} + \left(\frac{n}{log^2(1-\theta)(1-\theta)^2}\right) + \frac{n}{\log(1-\theta)(1-\theta)^2}$$

Exemplo: (continuação)

- Os valores do algoritmo serão atualizados da seguinte maneira

$$\theta_k = \theta_{k-1} - \frac{U(\theta_0)}{U'(\theta_0)} .$$

ou seja

$$\theta_k = \theta_{k-1} - \frac{\frac{\sum_i y_i}{\theta} + \frac{n}{\log(1-\theta)} \left(\frac{1}{1-\theta}\right)}{-\frac{\sum_i y_i}{\theta^2} + \left(\frac{n}{log^2(1-\theta)(1-\theta)^2}\right) + \frac{n}{\log(1-\theta)(1-\theta)^2}} .$$

Exemplo: (continuação)

- Suponha que observamos a amostra

$$\{1, 1, 1, 1, 1, 1, 2, 2, 2, 3\} .$$

- A função a seguir atualiza os valores do algoritmo

```
nr <- function(x)
{
 x.new<- x-(15/x+10/((1-x)*log(1-x)))/(-15/x^2+1
(((1-x)^2) *log(1-x)) +
 10/(((1-x)^2) *(log(1-x))^2))
 x.new
}
```

Exemplo: (continuação)

- Vamos definir o valor inicial $\theta_0 = 2/3$.
- O comando a seguir atualiza os valores até que o erro seja menor que 10^{-8}.

```
eps<-0.00000001
y.old<-2/3
delta<-1
while(delta>eps)
{
 y.new<- nr(y.old)
 delta<-sqrt((y.new-y.old)^2)
 y.old<-y.new
 print(y.old)
}
```

Exemplo: (continuação)

- A seguir encontram-se os valores obtidos pelo método

```
[1] 0.5497132
[1] 0.5336085
[1] 0.5335892
[1] 0.5335892
```

- O EMV de θ será

$$\hat{\theta}_{EMV} = 0.5335892\,.$$

Alguns problemas...

- Pode ser muito custoso calcular $U(\theta)$ e $U'(\theta)$.
- O método pode demorar a convergir.
- Pode oscilar muito.
- Uma alternativa:
  - substituir $U'(\theta)$ por $E(U'(\theta))$.
- Esse método é chamado **Mérodo Escore de Fisher**.

- Em muitos casos a $E(U'(\theta))$ é mais fácil de calcular do que $U'(\theta)$.
- O que é a $-E(U'(\theta))$? Informação de Fisher ($I_n(\theta_0)$)

$$-E(U'(\theta)) = -E\left(\frac{dU(\theta)}{d\theta}\right) = -E\left(\frac{d^2 l(\mathbf{y}, \theta)}{d\theta^2}\right)$$

- O método que usaremos para estimar os parâmetros do MLG será o Método Escore de Fisher.
- Veremos um exemplo de aplicação do método para encontrar o EMV.

Algoritmo Escore de Fisher

1. Escolha um valor inicial θ_0
2. Calcule
$$U(\theta_0) \quad E(U'(\theta_0)) \,.$$
3. Faça
$$\theta_1 = \theta_0 - \frac{U(\theta_0)}{E(U'(\theta_0))} \,.$$
4. Calcule $U(\theta_1)$.
5. Se $|U(\theta_1)| < 10^{-6}$ o algoritmo para.
6. Caso contrário, volta parao passo 2.

Pode ser escrito de outra maneira...

Algoritmo Escore de Fisher

1. Escolha um valor inicial θ_0
2. Calcule
$$U(\theta_0) \quad I_n(\theta_0)\,.$$
3. Faça
$$\theta_1 = \theta_0 + \frac{U(\theta_0)}{I_n(\theta_0))}\,.$$
4. Calcule $U(\theta_1)$.
5. Se $|U(\theta_1)| < 10^{-6}$ o algoritmo para.
6. Caso contrário, volta parao passo 2.

Exemplo:

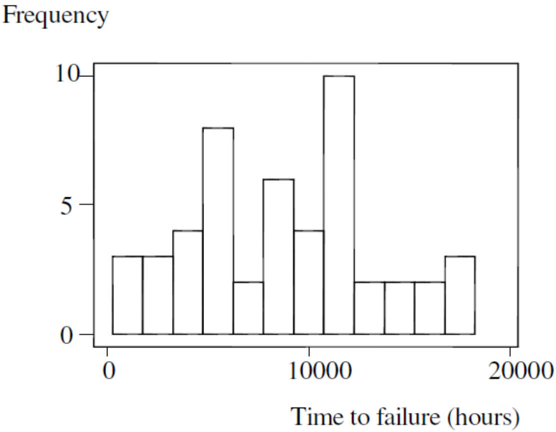
- Vasos de pressão são submetidos a um stress de 70%.
- Queremos analisar o tempo de falha desses vasos.
- A tabela mostra os dados coletados.

Table 4.1 *Lifetimes of pressure vessels.*

1051	4921	7886	10861	13520
1337	5445	8108	11026	13670
1389	5620	8546	11214	14110
1921	5817	8666	11362	14496
1942	5905	8831	11604	15395
2322	5956	9106	11608	16179
3629	6068	9711	11745	17092
4006	6121	9806	11762	17568
4012	6473	10205	11895	17568
4063	7501	10396	12044	

Exemplo: (continuação)

- A figura a seguir mostra o formato da distribuição dos dados.



Exemplo: (continuação)

- Podemos dizer que os dados tem distribuição normal?
- Aparentemente não.
- Uma distribuição muito usada nesse caso é a Weibull.
- Sua densidade é dada por

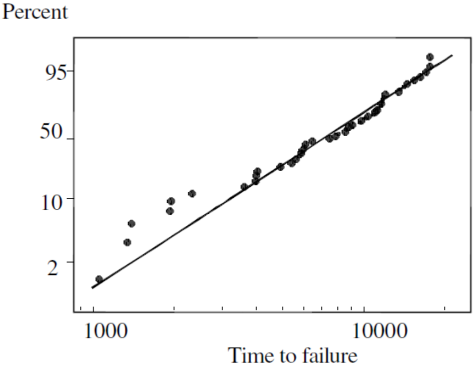
$$f(y, \lambda, \theta) = \frac{\lambda y^{\lambda-1}}{\theta^{\lambda}} \exp\left[-\left(\frac{y}{\theta}\right)^{\lambda}\right]$$

onde

  - $y > 0$ é o tempo de falha;
  - λ é parâmetro de forma da distribuição;
  - θ é parâmetro de escala da distribuição.

Exemplo: (continuação)

- Vamos verificar se a Weibull se ajusta bem a esses dados.
- A figura mostra o gráfico de probabilidade para $\lambda = 2$.



- Conclusão: a distribuição parece ser adequada, apesar de algumas discrepâncias.

Exemplo: (continuação)

- O parâmetro $\lambda = 2$ foi escolhido arbitrariamente.
- Queremos estimar apenas θ
- λ pode ser obtido por tentativa e erro.
- Na prática isso não é muito viável.
- Estamos apenas dando um exemplo.
- Existem métodos para estimar os dois parâmetros ao mesmo tempo.
- Vamos usar o método Escore de Fisher para estimar θ.

Exemplo: (continuação)

- Vamos primeiro encontrar a Função Escore $U(\theta)$.
- Vimos que a densidade é dada por

$$f(y, \lambda, \theta) = \frac{\lambda y^{\lambda-1}}{\theta^{\lambda}} \exp\left[-\left(\frac{y}{\theta}\right)^{\lambda}\right].$$

- Portanto a função de verossimilhança é dada por

$$f(\mathbf{y}, \lambda, \theta) = \prod_{i=1}^{n} \frac{\lambda y_i^{\lambda-1}}{\theta^{\lambda}} \exp\left[-\left(\frac{y_i}{\theta}\right)^{\lambda}\right].$$

- Tomando o logaritmo ficamos com

$$\log(f(\mathbf{y}, \lambda, \theta)) = \sum_{i=1}^{n} \log\left\{\frac{\lambda y_i^{\lambda-1}}{\theta^{\lambda}} \exp\left[-\left(\frac{y_i}{\theta}\right)^{\lambda}\right]\right\}.$$

Exemplo: (continuação)

$$= \sum_{i=1}^{n} \left(\log \left\{ \frac{\lambda y_i^{\lambda-1}}{\theta^{\lambda}} \right\} - \left(\frac{y_i}{\theta} \right)^{\lambda} \right)$$

$$= \sum_{i=1}^{n} \left(\log(\lambda) + (\lambda - 1) \log(y_i) - \lambda \log(\theta) - y_i^{\lambda} \theta^{-\lambda} \right)$$

- Derivando a log-verossimilhança com relação a θ

$$U(\theta) = \frac{d \log(f(\mathbf{y}, \theta))}{d\theta} = \sum_{i=1}^{n} \left(-\frac{\lambda}{\theta} + \lambda y_i^{\lambda} \theta^{-\lambda-1} \right)$$

$$= \sum_{i=1}^{n} \left(-\frac{\lambda}{\theta} + \frac{\lambda y_i^{\lambda}}{\theta^{\lambda+1}} \right)$$

Exemplo: (continuação)

- O EMV de θ é o valor que satisfaz a equação

$$U(\theta) = \sum_{i=1}^{n} \left(-\frac{\lambda}{\theta} + \frac{\lambda y_i^{\lambda}}{\theta^{\lambda+1}} \right) = 0 \,.$$

- Como obter o EMV nesse caso?
- Conseguimos isolar θ na equação? Não.
- Precisamos recorrer a métodos numéricos.
- Vamos usar o método Escore de Fisher.

Exemplo: (continuação)

- Vamos encontrar a derivada segunda da Função Escore $U(\theta)$.
- Vimos que

$$U(\theta) = \sum_{i=1}^{n} \left(-\frac{\lambda}{\theta} + \lambda y_i^{\lambda} \theta^{-(\lambda+1)} \right)$$

$$\frac{dU(\theta)}{d\theta} = \sum_{i=1}^{n} \left(\frac{\lambda}{\theta^2} + \lambda(\lambda+1) \frac{y_i^{\lambda}}{\theta^{\lambda+2}} \right)$$

$$= n\frac{\lambda}{\theta^2} + \frac{\lambda(\lambda+1)}{\theta^{\lambda+2}} \sum_i y_i^{\lambda}$$

Exemplo: (continuação)

- Fixando $\lambda = 2$, ficamos com

$$\frac{dU(\theta)}{d\theta} = n\frac{2}{\theta^2} + \frac{2(2+1)}{\theta^{2+2}} \sum_i y_i^2$$

- Para encontrarmos o valor de $E(U'(\theta)$ vamos utilizar um resultado.
- Se a função de densidade pode ser escrita na forma

$$f(y, \theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]$$

então

$$E(U'(\theta)) = c''(\theta) - \frac{b''(\theta)c'(\theta)}{b'(\theta)}\ .$$

Exemplo: (continuação)

- Vamos reescrever a função Weibull no formato da família exponencial.
- Temos que

$$f(y,\lambda,\theta) = \frac{\lambda y^{\lambda-1}}{\theta^{\lambda}} \exp\left[-\left(\frac{y}{\theta}\right)^{\lambda}\right]$$

$$= \exp\left\{\log(\lambda) + (\lambda-1)\log(y) - \lambda\log(\theta) - \left(\frac{y}{\theta}\right)^{\lambda}\right\}$$

$$= \exp\left\{y^{\lambda}\underbrace{(-\theta^{\lambda})}_{b(\theta)}\underbrace{-\lambda\log(\theta)}_{c(\theta)}\underbrace{+(\lambda-1)\log(y)+\log(\lambda)}_{d(y)}\right\}$$

Exemplo: (continuação)

- Portanto

$$b(\theta) = -\theta^{-\lambda} \qquad c(\theta) = -\lambda \log(\theta)$$

$$b'(\theta) = \lambda\theta^{-\lambda-1} \quad b''(\theta) = -\lambda(\lambda+1)\theta^{-\lambda-2}$$

$$c'(\theta) = -\frac{\lambda}{\theta} \quad c''(\theta) = \frac{\lambda}{\theta^2}$$

- Temos então que

$$E(U'(\theta)) = c''(\theta) - \frac{b''(\theta)c'(\theta)}{b'(\theta)}.$$

$$= \frac{\lambda}{\theta^2} - \frac{-\lambda(\lambda+1)\theta^{-\lambda-2}(-\frac{\lambda}{\theta})}{\lambda\theta^{-\lambda-1}}$$

Exemplo: (continuação)

$$= \frac{\lambda}{\theta^2} - \frac{\lambda(\lambda+1)}{\theta^2}$$

$$= \frac{\lambda - \lambda^2 - \lambda}{\theta^2} = \frac{\lambda^2}{\theta^2}$$

- Como temos uma amostra de tamanho n ficamos com

$$E(U'(\theta)) = n\left(c''(\theta) - \frac{b''(\theta)c'(\theta)}{b'(\theta)}\right) = n\frac{\lambda^2}{\theta^2}\,.$$

Exemplo: (continuação)

- A atualização dos valores do algoritmo é dada por

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

$$\theta_k = \theta_{k-1} - \frac{\sum_{i=1}^{n} \left(-\frac{\lambda}{\theta} + \lambda y_i^{\lambda} \theta^{-(\lambda+1)}\right)}{n\frac{\lambda^2}{\theta^2}}$$

Exemplo: (continuação)

- A tabela a seguir mostra os passos de iteração do algoritmo

Iteration	1	2	3	4
θ	8805.9	9633.9	9876.4	9892.1
$U \times 10^6$	2915.10	552.80	31.78	0.21
$U' \times 10^6$	-3.52	-2.28	-2.02	-2.00
$\mathrm{E}(U') \times 10^6$	-2.53	-2.11	-2.01	-2.00
U/U'	-827.98	-242.46	-15.73	-0.105
$U/\mathrm{E}(U')$	-1152.21	-261.99	-15.81	-0.105

Exemplo: (continuação)

- O valor inicial utilizado foi a média dos dados

$$\theta_0 = 8805,9\,.$$

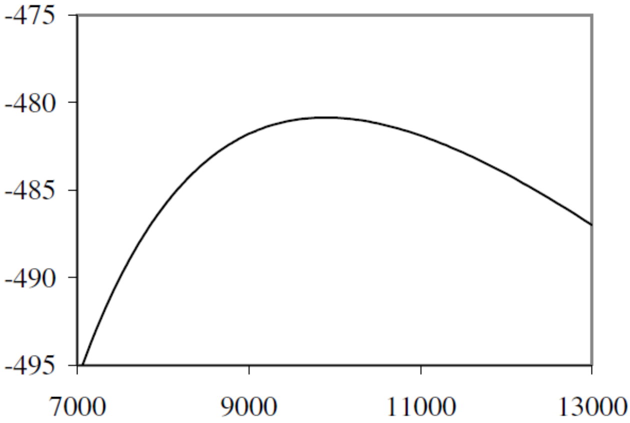
- A tabela mostra que

$$U'(\theta) \approx E(U'(\theta))$$

portanto, poderíamos usar Newton Raphson ou Escore de Fisher.

Exemplo: (continuação)

- A figura a seguir mostra função de log-verossimilhança.



- O EMV de θ está em torno de quanto? 980,00