Obtenção de estimativas de máxima verossimilhança via algoritmo Escore de Fisher

Prof. Caio Azevedo

9 de maio de 2011

- Em muitos situações, a verossimilhança é uma função contínua dos parâmetros de interesse.
- Nestes casos, é possível obter os máximos globais através de métodos tradicionais de maximização.
- Isto equivale à resolver sistemas de equações oriundos das derivadas parciais de primeira ordem da log-verossimilhança.
- Problema: em alguns casos não é possível obter soluções analíticas para esses sistemas de equações.

- Seja $X_1, ..., X_n$ uma amostra aleatória de $X \sim gama(r, \theta)$, θ conhecido
- Verossimilhança

$$L(r) = e^{-\sum_{i=1}^{n} \frac{x_i}{\theta}} \prod_{i=1}^{n} x_i^{r-1} \frac{1}{(\Gamma(r))^n \theta^{nr}}$$

- Log-verossimilhança

$$l(r) = -n \ln(\Gamma(r)) - nr \ln(\theta) - \sum_{i=1}^{n} \frac{x_i}{\theta} + (r-1) \sum_{i=1}^{n} \ln(x_i)$$

- Função escore

$$S(r) = -n\frac{\Gamma'(r)}{\Gamma(r)} - n\ln\theta + \sum_{i=1}^{n}\ln(x_i) = -n\Psi(r) - n\ln\theta + \sum_{i=1}^{n}\ln x_i \quad (1)$$

- Função Hessiana

$$H(r) = -\frac{n}{(\Gamma(r))^2}\left[\Gamma''(r)\Gamma'(r) - (\Gamma'(r))^2\right] = -n\Psi'(r)$$

em que $\Psi(r)$ e $\Psi'(r)$ são, respectivamente, a função digama e a função trigama.

- Informação de Fisher

$$I(\theta) = \frac{n}{(\Gamma(r))^2}\left[\Gamma''(r)\Gamma'(r) - (\Gamma'(r))^2\right] = n\Psi'(r)$$

- A equação $S(\widetilde{r}) = 0$ não possui solução explícita.
- Solução: utilização de métodos numéricos para obtenção de raízes de equações (não-lineares).
- Algoritmo de Newton-Raphson.
- Versão estocástica : Algoritmo escore de Fisher.
- Suporte teórico: expansão em séries de Taylor da função Escore.

- É um método iterativo.
- Inicializa-se com um "chute" inicial (de preferências um valor próximo do verdadeiro valor do parâmetros ou dos parâmetros).
- Com o chute inicial, gera-se uma nova estimativa para o parâmetro.
- Verifica-se se o critério de parada foi satisfeito.
- Se sim, termina-se o processo, caso contrário, gera-se um novo valor para o parâmetro com base na estimativa anterior.
- Repete-se os dois passos anteriores, até que o critério de parada seja alcançado.

- Seja $\widetilde{\boldsymbol{\theta}}^{(t)}$ uma estimativa para $\boldsymbol{\theta}$ obtido na iteração t.
- Obtem-se uma estimativa atualizada de $\boldsymbol{\theta}$, digamos $\widetilde{\boldsymbol{\theta}}^{(t+1)}$, através de

$$\widetilde{\boldsymbol{\theta}}^{(t+1)} = \widetilde{\boldsymbol{\theta}}^{(t)} + \mathbf{I}\left(\widetilde{\boldsymbol{\theta}}^{(t)}\right)^{-1}\mathbf{S}\left(\widetilde{\boldsymbol{\theta}}^{(t)}\right)$$

- Critério de parada: por exemplo $||\widetilde{\boldsymbol{\theta}}^{(t+1)} - \widetilde{\boldsymbol{\theta}}^{(t)}|| < \epsilon, \epsilon > 0$

- Distribuição assintótica de $\widehat{r}$ (emv de r). Para n suficientemente grande, temos que

$$\widehat{r} \approx N\left(r, \frac{1}{n\Psi'(r)}\right)$$

- Erro-padrão assintótico de $\widehat{r}$, $EP_A(\widehat{r}) = \sqrt{\frac{1}{n\Psi'(r)}}$.