

Plano Aulas 31 e 32

Markus Stein

02 July 2019

Otimização

Considere o problema de otimização da função h

$$\max_{\theta \in \Theta} h(\theta).$$

- Usualmente $h(\cdot)$ é uma **função de verossimilhança** $L(\theta)$ ou uma **distribuição *a posteriori*** $\pi(\theta|Y)$.
- Note que para o problema acima $\max_{\theta \in \Theta} h(\theta) = \min_{\theta \in \Theta} -h(\theta)$.

Exemplo Modelo Normal - Modelo Linear

- **Exemplo 1:** Questão 3 da prova 2: Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tal que $Y_i \sim \text{Normal}(\beta x_i, \sigma^2)$. Assumimos assim que $Y = \beta x + \epsilon$, $\epsilon \sim \text{Normal}(0, \sigma^2)$.
- Qual a variável resposta?
- Qual a distribuição de probabilidade da variável resposta?
- Como estimar os parâmetros $\theta = (\beta, \sigma^2)$?

$$\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2) = \arg \max_{\theta} L(\beta, \sigma^2).$$

Exemplo Modelo Logístico - Modelo não linear

- **Exemplo 2:** Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tal que $Y_i \sim \text{Bernoulli}(\pi_i, \sigma^2)$, onde $\pi_i = \exp(\alpha + \beta x_i) / (1 + \exp(\alpha + \beta x_i))$.
- Como estimar os parâmetros $\theta = (\alpha, \beta)$?

Métodos Gradiente

- **Gradiente:** o gradiente de uma função $h(\theta)$ é dado pela sua derivada (ou vetor de derivadas):
 - no caso θ unidimensional - $\frac{\partial}{\partial \theta} h(\theta)$;
 - no caso θ multidimensional - $\nabla h(\theta) = \frac{\partial}{\partial \theta} h(\theta) = \left(\frac{\partial}{\partial \theta_1} h(\theta), \dots, \frac{\partial}{\partial \theta_k} h(\theta) \right)$.
- Relembrando **expansão em série de Taylor** - (Teorema 5.5.21, Casella e Berger): Para uma variável aleatória X tal que $E(X) = \mu \neq 0$, uma aproximação para estimar $g(\mu)$ é dada por $g(X) \approx g(\mu) + g'(\mu)(X - \mu)$, onde $g'(\mu) = \frac{\partial}{\partial \mu} g(\mu)$.

Newthon-Raphson (Tanner, pg. 14)

Para **EMV**, na expressão acima defina $X = \hat{\theta}$, $\mu = \theta$ e escolha a função escore $g(\theta) = U(\theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \frac{\partial}{\partial \theta} \log L(\theta)$ para a expansão. Então, expandindo $U(\hat{\theta})$ ao redor de θ temos

$$0 = U(\hat{\theta}) \approx U(\theta) + (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} U(\theta) \iff \hat{\theta} \approx \theta + U(\theta) [J(\theta)]^{-1},$$

onde $J(\theta) = -\frac{\partial}{\partial \theta} U(\theta) = -\frac{\partial^2}{\partial \theta^2} \ell(\theta)$.

- Note que $J(\theta)$ é a **informação de Fisher observada**.
- Newthon-Raphson pode não ser estável se ocorre separação, pequenas amostras, ...

Scoring de Fisher

$$\hat{\theta} \approx \theta + U(\theta) [I(\theta)]^{-1},$$

- Usa a informação de Fisher (esperada) ao invés da observada $I(\theta) = E[J(\theta)] = E\left[-\frac{\partial^2}{\partial \theta^2} \ell(\theta)\right]$.

Otimização com restrição

Problemas em otimização: Missing data x censura x causal? - complex sampling x mixed models

Tarefa: Fazer trabalho final para entregar e apresentar dia 11/07.

Referências

Tanner (1996) Tools for Statistical Inference.

Robert e Casella (2010) Introduction to MOnTe Carlo Methods with R

James, Witten, Hastie e Tibishirani (2010) An Introduction to Statistical Learning with Applications in R
<http://www-bcf.usc.edu/~gareth/ISL/>

Curso UFPR - Métodos Computacionais para Inferência Estatística <http://www.leg.ufpr.br/~wagner/MCIE/index.html> Apostila Justiniano - <http://www.leg.ufpr.br/lib/exe/fetch.php/pessoais:mcie-sinape-v12.pdf>