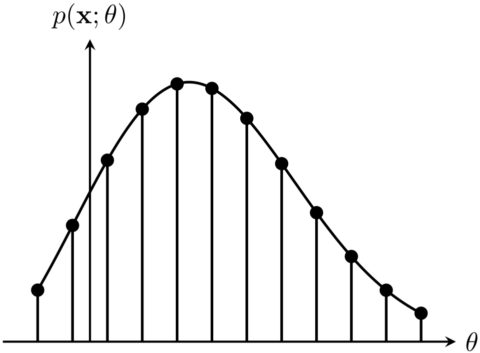
Statistical signal processing
5CTA0

Estimation theory - numerical methods

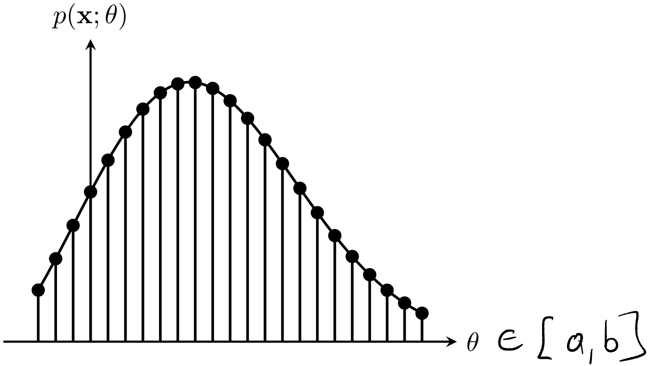
Numerical methods

- Many problems cannot be solved analytically
- Resort to numerical methods
  - Grid search
  - Iterative methods based on linearization of optimization problem

Grid search

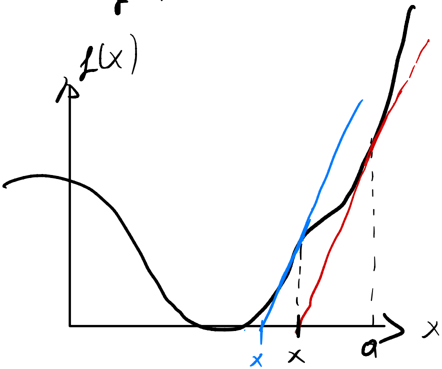


Grid search



Newton-Raphson method

$$f(x) = 0$$



$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

$$f'(a) \approx \frac{\overbrace{f(x)}^{a} - f(a)}{x - a}$$

$$x = a - \frac{f(a)}{f'(a)}$$

Newton-Raphson for MLE - scalar parameter

$$\underbrace{\frac{\partial \ln p(\underline{x},\theta)}{\partial \theta}}_{f(x)} = 0$$

- Newton Raphson for MLE:

$$\theta_{m+1} = \theta_m - \Big(\underbrace{\frac{\partial^2 \ln p(\mathbf{x};\theta_m)}{\partial \theta^2}}_{f'(\theta)}\Big)^{-1} \frac{\partial \ln p(\mathbf{x};\theta_m)}{\partial \theta}$$

- Method of scoring:

$$\theta_{m+1} = \theta_m + \mathcal{I}^{-1}(\theta_m)\frac{\partial \ln p(\mathbf{x};\theta_m)}{\partial \theta}$$

$$-E\left[\frac{\partial^2}{\partial \theta} \ln p(x;\theta)\right]$$

$$\frac{\partial^2 \ln p(\underline{x};\theta)}{\partial \theta^2} = \sum_{n=0}^{N-1} \frac{\partial \ln p(x_n;\theta)}{\partial \theta^2} = \overbrace{N \underbrace{\frac{1}{N}\sum_{n=0}^{N-1} \frac{\partial \ln p(x_n;\theta)}{\partial \theta^2}}_{\to E\left[\frac{\partial^2 \ln p(x_n;\theta)}{\partial \theta^2}\right] = i(\theta)}}^{I(\theta)}$$

Newton-Raphson for MLE - vector parameter

$J(\theta) = \sum_{n=0}^{N-1} (x_n - s_n(\theta))^2$

- Newton-Raphson iteration:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m - \mathbf{H}^{-1}(\boldsymbol{\theta}_m) \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\mathbf{x}; \boldsymbol{\theta}_m)$$

- Hessian matrix:

$$\mathbf{H}(\boldsymbol{\theta})_{i,j} = \begin{bmatrix} \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial^2 \theta_0^2} & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_0 \partial \theta_1} & \cdots & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_0 \partial \theta_{K-1}} \\ \frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_1 \partial \theta_0} & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \partial \theta_1^2} & \cdots & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_1 \partial \theta_{K-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_{K-1} \partial \theta_0} & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_{K-1} \partial \theta_1} & \cdots & \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \theta_{K-1^2}} \end{bmatrix}$$

- Method of scoring:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m + \mathbf{I}^{-1}(\boldsymbol{\theta}_m) \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\mathbf{x}; \boldsymbol{\theta}_m)$$

Gauss-Newton method for LS - scalar parameter

- Scalar parameter:

$s_n(\theta) = s_n(\theta_0) + \frac{\partial}{\partial\theta} s_n(\theta_0)(\theta - \theta_0)$

$0 \le n \le N-1$

$$J(\theta) = \sum_{n=0}^{N-1} (x_n - s_n(\theta))^2$$
$$\approx \sum_{n=0}^{N-1} \left(x_n - \left(s_n(\theta_0) + \frac{\partial s_n(\theta_0)}{\partial\theta}(\theta - \theta_0) \right) \right)^2$$
$$= \sum_{n=0}^{N-1} \left(x_n - s_n(\theta_0) + \frac{\partial s_n(\theta_0)}{\partial\theta}\theta_0 - \frac{\partial s_n(\theta_0)}{\partial\theta}\theta \right)^2$$
$$= (\mathbf{x} - \mathbf{s}(\theta_0) + \mathbf{H}(\theta_0)\theta_0 - \mathbf{H}(\theta_0)\theta)^T (\mathbf{x} - \mathbf{s}(\theta_0) + \mathbf{H}(\theta_0)\theta_0 - \mathbf{H}(\theta_0)\theta)$$

with

$$\mathbf{H}(\theta) = \begin{bmatrix} \frac{\partial s_0(\theta)}{\partial\theta} & \frac{\partial s_1(\theta)}{\partial\theta} & \dots & \frac{\partial s_{N-1}(\theta)}{\partial\theta} \end{bmatrix}^T$$

Gauss-Newton method for LS - scalar parameter

- Known parts:

$$\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{s}(\theta_0) + \mathbf{H}(\theta_0)\theta_0$$

- Approximated cost function:

$$J(\theta) \approx (\tilde{\mathbf{x}} - \mathbf{H}(\theta_0)\theta)^T (\tilde{\mathbf{x}} - \mathbf{H}(\theta_0)\theta)$$

$$\hat{\theta}_{LS} = (H^T H)^{-1} H^T \underline{x}$$

- Parameter estimate:

$$\begin{aligned}\hat{\theta} &= \left(\mathbf{H}^T(\theta_0)\mathbf{H}(\theta_0)\right)^{-1} \mathbf{H}^T(\theta_0)\tilde{\mathbf{x}} \\ &= \theta_0 + \left(\mathbf{H}^T(\theta_0)\mathbf{H}(\theta_0)\right)^{-1} \mathbf{H}^T(\theta_0)\left(\mathbf{x} - \mathbf{s}(\theta_0)\right)\end{aligned}$$

- Iteration:

$$\theta_{m+1} = \theta_m + \left(\mathbf{H}^T(\theta_m)\mathbf{H}(\theta_m)\right)^{-1} \mathbf{H}^T(\theta_m)\left(\mathbf{x} - \mathbf{s}(\theta_m)\right)$$

Gauss-Newton method for LS - vector parameter

- Iteration:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m + \left(\mathbf{H}^T(\boldsymbol{\theta}_m)\mathbf{H}(\boldsymbol{\theta}_m)\right)^{-1} \mathbf{H}^T(\boldsymbol{\theta}_m)\left(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}_m)\right)$$

- Jacobian matrix:

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial s_0(\boldsymbol{\theta})}{\partial \theta_0} & \frac{\partial s_0(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial s_0(\boldsymbol{\theta})}{\partial \theta_{K-1}} \\ \frac{\partial s_1(\boldsymbol{\theta})}{\partial \theta_0} & \frac{\partial s_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial s_1(\boldsymbol{\theta})}{\partial \theta_{K-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial s_{N-1}(\boldsymbol{\theta})}{\partial \theta_0} & \frac{\partial s_{N-1}(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial s_{N-1}(\boldsymbol{\theta})}{\partial \theta_{K-1}} \end{bmatrix}$$