Fisher Information

- Consider $\vec{y}$ + $\vec{\theta}$, with joint pdf $P_y(\vec{y}|\vec{\theta})$.
  - Let $\lambda(\vec{y}|\theta) = \ln(P_y(\vec{y}|\theta))$ be the log-pdf.
  - Assume that P_y is differentiable

- The Fisher Information, $\mathcal{F}(\vec{\theta})$, is defined by

$$\mathcal{F}(\vec{\theta}) = Var[S(\vec{y}|\vec{\theta})], \quad S(\vec{y}|\theta) = \frac{\partial}{\partial\theta}\lambda(\vec{y}|\theta)$$

where $S(\vec{y}|\vec{\theta})$ is the score.

★ $\mathcal{F}(\vec{\theta})$ measures the information in which $\vec{y}$ describes $\vec{\theta}$. ★

- To derive this: note that

$$\int_y P_y(y|\theta)\,dy = 1$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int_y P(y|\theta)\,dy = 0$$

Note

$$\Rightarrow \frac{\partial}{\partial\theta}\int_y P_y(y|\theta)\,dy = \int_y \frac{\frac{\partial}{\partial\theta}P_y(y|\theta)}{P_y(y|\theta)} \cdot P_y(y|\theta)\,dy = 0$$

$$= \int_y \frac{\partial}{\partial\theta}\ln(P_y(\vec{y}|\vec{\theta})) \cdot P_y(\vec{y}|\vec{\theta})\,dy$$

(★) $$= \int \frac{\partial}{\partial\theta}\lambda(\vec{y}|\vec{\theta})\,P_y(\vec{y}|\vec{\theta})\,dy = 0$$

$$\ldots [S(y|\theta)] = Var\left[\frac{\partial}{\partial\theta}\lambda(y|\theta)\right]$$

- Looking at $\mathcal{F}(\theta) = Var[S(y|\theta)] = Var\left[\frac{\partial}{\partial\theta}\lambda(y|\theta)\right]$

$$= E[S^2] - (E[S])^2$$

- Note $E[S] = E\left[\frac{\partial}{\partial\theta}\lambda(\vec{y}|\theta)\right] = \int_y \frac{\partial}{\partial\theta}\lambda(\vec{y}|\theta)\,dy = 0$ (from $\ast$)

$$\Rightarrow \mathcal{F}(\theta) = E[S^2] - \cancelto{0}{(E[S])^2} = E[S^2]$$

$$\boxed{\Rightarrow \mathcal{F}(\theta) = -E\left[\frac{\partial^2\lambda}{\partial\theta^2}\right]}$$

- The Fisher information is related to the negative second derivative of the log-likelihood.

- For a finite sample, $\vec{y} \in \mathbb{R}^n$, the Fisher information matrix (FIM) is

$$\underset{\sim}{F} = -E[\nabla^2_\theta \lambda] = E[\nabla_\theta \lambda (\nabla_\theta\lambda)^T]$$

Ex) Suppose $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

- $\lambda(\vec{y}|\theta) \propto -\frac{1}{2\sigma^2} SS(\theta) = \frac{-1}{2\sigma}\sum_{i=1}^{N_x}(y_i - f(x_i;\theta))^2$

$\Rightarrow \frac{\partial^2}{\partial\theta^2}\lambda = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(\frac{-1}{2\sigma^2}SS(\vec{\theta})\right)$, $i,j = 1,\dots,p$

$$\frac{1}{\ } \left[\frac{\partial}{\partial\theta} \sum^{N_y} (y_x - f(x_x;\vec{\theta}))\frac{\partial f}{\partial\theta_j}\right]$$

$$= -\frac{1}{\sigma^2}\left[\frac{\partial}{\partial\theta_i}\sum_{q=1}^{N_y}\left(y_q - f(\underset{\sim}{x}_q;\vec{\theta})\right)\frac{\partial f}{\partial\theta_j}\right]$$

$$= -\frac{1}{\sigma^2}\left[\sum_{q=1}^{N_y}\left(y_q - f(\underset{\sim}{x}_q;\vec{\theta})\right)\frac{\partial^2 f}{\partial\theta_i\partial\theta_j} - \frac{\partial f}{\partial\theta_i}\frac{\partial f}{\partial\theta_j}\right]$$

- Note

$$E\left[\sum_{q=1}^{N_y}\left(y_q - f(x_q;\theta)\right)\right] = 0, \quad \text{because } E[\varepsilon] = 0$$

$$\Rightarrow \underset{\sim}{F} = -E\left[\frac{\partial^2 \lambda}{\partial\theta^2}\right] = \frac{1}{\sigma^2}E\left[-\frac{\partial f}{\partial\theta_i}\frac{\partial f}{\partial\theta_j}\right] = -\frac{1}{\sigma^2}\frac{\partial f}{\partial\theta_i}\frac{\partial f}{\partial\theta_j} = -\frac{1}{\sigma^2}\underset{\sim}{S}^T\underset{\sim}{S}$$

$$\text{where } \underset{\sim}{S} = \begin{bmatrix} \frac{\partial f}{\partial\theta_1}(x_1) & \cdots & \frac{\partial f}{\partial\theta_p}(x_1) \\ \vdots & & \\ \frac{\partial f}{\partial\theta_1}(x_{N_y}) & \cdots & \frac{\partial f}{\partial\theta_p}(x_{N_y}) \end{bmatrix}$$