# Fisher Information

(1) Def:

For $n$ samples $\{X_i\}_1^n \sim f(\vec{x}|\theta)$. $l(\theta|\vec{x}) = \log f(\vec{x}|\theta)$

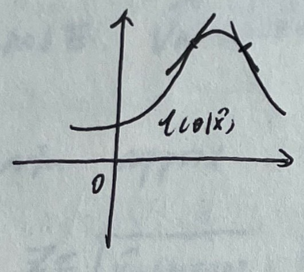
$$I_{\vec{x}}(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta} l(\theta|\vec{x})\right)^2\right) = \left\|\frac{\partial}{\partial\theta} l(\theta|\vec{x})\right\|_{L^2(f)}$$

Remark: i) Under regular condition: $I = -E\left(\frac{\partial^2}{\partial\theta^2} l(\theta|\vec{x})\right)$

ii) Under regular condition: $E_\theta\left(\frac{\partial}{\partial\theta} l(\theta|\vec{x})\right) = 0$

$\therefore I_{\vec{x}}(\theta) = Var\left(\frac{\partial}{\partial\theta} l(\theta|\vec{x})\right)$

Interpretation:




It's average weight of slope of $l(\theta|\vec{x})$ in $L^2$ norm.

It quantizes how rapidly $l(\theta|\vec{x})$ changes. If $l(\theta|x)$ differs a lot in the space with different $\theta$. Then $\{X_i\}_1^n$ definitely can provide lots of information.

(2) Properties:

① Thm. $\{X_k\}_1^n$ i.i.d. Then $I_{\vec{x}}(\theta) = \sum_1^n I_{X_i}(\theta) = n I_{X_i}(\theta)$

Remark: It means indept. r.v.'s offer the information that won't overlap.

Note that if $Cor(X, Y) = 1$. then $Y = aX + b$. $\therefore I_{(X,Y)}(\theta) = I_X(\theta)$

Pf: If $X \sim f(x|\theta)$. Then $(X, Y) \sim f(x+\lambda|\theta)$ for some $\lambda$. translation.

② $I_{T(\vec{x})}(\theta) = I_{\vec{x}}(\theta)$. where $T(\vec{x})$ is the sufficient statistics for $\theta$.

Remark: Note that $I_{\vec{x}}(\theta)$ only concerns on some specific parameter $\theta$. $T(\vec{x})$ retains all information about $\theta$.

Pf: Note that $f(\vec{x}|\theta) = g(T(\vec{x})|\theta)h(\vec{x})$

$\therefore \frac{\partial}{\partial\theta} \log f(\vec{x}|\theta) = \frac{\partial}{\partial\theta} \log g(T(\vec{x})|\theta)$

③ Under reparametrization:

$I_{\vec{x}}(\tau(\theta)) = E_\theta((\frac{\partial}{\partial\tau(\theta)} l(\theta|\vec{x}))^2) = \frac{I_{\vec{x}}(\theta)}{\tau'(\theta)}$

Pf: Note that $\tau(\theta)$ influences the slope of $l(\theta|\vec{x})$ by a dilation $\tau'(\theta)^{-1}$.

(3) Multidimensional case:

$\vec{\theta} = (\theta_1, \theta_2 \cdots \theta_K)^T$. $I_{\vec{x}}(\vec{\theta})$ is a $K\times K$ matrix.

where $I_{\vec{x}}(\vec{\theta})(i,j) = E(\frac{\partial}{\partial\theta_i} l(\vec{\theta}|\vec{x}) \frac{\partial}{\partial\theta_j} l(\vec{\theta}|\vec{x}))$

$(= -E(\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\vec{\theta}|\vec{x}))$ under regular condition)