Fisher Information
Stat 543 Spring 2005

Definition 1 *We will say that the model for X specified by $f(x|\theta)$ (either a probability mass function or a probability density) is FI (Fisher Information) Regular at $\theta_0 \in \Theta \subset \mathcal{R}^k$ provided there is an open neighborhood of θ_0, say $\mathcal{O}$, such that*

i) $f(x|\theta) > 0 \, \forall x$ *and* $\forall \theta \in \mathcal{O}$,

ii) $\forall x$, $f(x|\theta)$ *has first order partials at* θ_0, *and*

iii) $0 = \sum_x \frac{\partial}{\partial \theta_i} f(x|\theta)\Big|_{\theta=\theta_0}$ *(or* $0 = \int \frac{\partial}{\partial \theta_i} f(x|\theta)\Big|_{\theta=\theta_0} dx$*)* .

Definition 2 *If the model for X is FI regular at θ_0 and*

$$E_{\theta_0} \left(\frac{\partial}{\partial \theta_i} \ln f(X|\theta)\Big|_{\theta=\theta_0} \right)^2 < \infty \; \forall i \ ,$$

then the $k \times k$ matrix

$$I(\theta_0) = \left(E_{\theta_0} \frac{\partial}{\partial \theta_i} \ln f(X|\theta)\Big|_{\theta=\theta_0} \frac{\partial}{\partial \theta_j} \ln f(X|\theta)\Big|_{\theta=\theta_0} \right)$$

is called the Fisher Information Matrix at θ_0.

Theorem 3 *Suppose the model for X is FI Regular at $\theta_0 \in \Theta \subset \mathcal{R}^k$ and that $I(\theta_0)$ exists and is nonsingular. If $\forall x$ $f(x|\theta)$ has continuous second order partials in the neighborhood $\mathcal{O}$, and $\forall i, j$*

$$0 = \sum_x \frac{\partial}{\partial \theta_i} f(x|\theta)\Big|_{\theta=\theta_0} \text{ and } 0 = \sum_x \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x|\theta)\Big|_{\theta=\theta_0} \text{ in the discrete case}$$

or

$$0 = \int \frac{\partial}{\partial \theta_i} f(x|\theta)\Big|_{\theta=\theta_0} dx \text{ and } 0 = \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x|\theta)\Big|_{\theta=\theta_0} dx \text{ in the continuous case,}$$

then

$$I(\theta_0) = -E_{\theta_0} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(X|\theta)\Big|_{\theta=\theta_0} \right) \quad .$$

Notation 4 *In an iid model, i.e. where $X = (X_1, ..., X_n)$, let $I_1(\theta_0)$ be the Fisher Information Matrix at θ_0 for a single observation. That is, with marginal distribution specified by $f(x|\theta)$,*

$$I_1(\theta_0) = \left(E_{\theta_0} \frac{\partial}{\partial \theta_i} \ln f(X_1|\theta)\Big|_{\theta=\theta_0} \frac{\partial}{\partial \theta_j} \ln f(X_1|\theta)\Big|_{\theta=\theta_0} \right) \quad .$$

Theorem 5 *Suppose that $X_1, X_2, ..., X_n$ are iid random vectors and $f(x|\theta)$ is either the marginal probability mass function or the marginal probability density. If the model for X_1 is FI regular at θ_0 and with $X = (X_1, ..., X_n)$, $I(\theta_0)$ is the Fisher Information Matrix for X at θ_0,*

$$I(\theta_0) = n I_1(\theta_0) \quad .$$