# Stat 215B (Spring 2005): Fisher Information


GSI: Victor Panaretos

victor@stat.berkeley.edu


March 18 2005

Let $\mathcal{P} = \{P(\cdot|\widetilde{\theta})\}_{\widetilde{\theta}\in\Theta}$ be a parametric family of distributions, with corresponding pdfs/pmfs $\{f(\cdot|\widetilde{\theta})\}_{\widetilde{\theta}\in\Theta}$. Suppose that we have a random sample $X_1, ..., X_n$ from a distribution that is a member of that family, say $P(\cdot|\widetilde{\theta}^{(0)})$, with pdf/pmf $f(\cdot|\widetilde{\theta}^{(0)})$. Let $\ell(\widetilde{\theta})$ be the log-likelihood as a function of the parameter $\widetilde{\theta} = (\theta_1, \ldots \theta_p)^T$ (the true parameter is $\widetilde{\theta}^{(0)} = (\theta_1^{(0)}, \ldots \theta_p^{(0)})^T$).

**Fisher Information Matrix**

The Fisher information matrix is defined as:

$$I(\widetilde{\theta}^{(0)}) = -\begin{bmatrix} \mathbb{E}\left\{ \frac{\partial^2 \ell(\widetilde{\theta})}{\partial \theta_1^2}\Big|_{\widetilde{\theta}^{(0)}} \right\} & \cdots & \mathbb{E}\left\{ \frac{\partial^2 \ell(\widetilde{\theta})}{\partial \theta_1 \partial \theta_p}\Big|_{\widetilde{\theta}^{(0)}} \right\} \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left\{ \frac{\partial^2 \ell(\widetilde{\theta})}{\partial \theta_p \partial \theta_1}\Big|_{\widetilde{\theta}^{(0)}} \right\} & \cdots & \mathbb{E}\left\{ \frac{\partial^2 \ell(\widetilde{\theta})}{\partial \theta_p^2}\Big|_{\widetilde{\theta}^{(0)}} \right\} \end{bmatrix}$$

Notice that since $\{X_i\}_{i=1}^n$ are iid, we have that $I(\widetilde{\theta}^{(0)}) = nF(\widetilde{\theta}^{(0)})$, where $F_{ij}(\widetilde{\theta}^{(0)}) = -\mathbb{E}\left\{ \frac{\partial^2 \log f(X_k|\widetilde{\theta})}{\partial \theta_i \partial \theta_j}\Big|_{\widetilde{\theta}^{(0)}} \right\}$, for any $k \in \{1, ...n\}$.

**A useful relationship**

It can be shown that:

$$-\mathbb{E}\left\{ \frac{\partial^2 \ell(\widetilde{\theta})}{\partial \theta_i \partial \theta_j} \right\} = \mathbb{E}\left\{ \frac{\partial \ell(\widetilde{\theta})}{\partial \theta_i} \frac{\partial \ell(\widetilde{\theta})}{\partial \theta_j} \right\}$$

**Asymptotic Normality of the MLE**

Suppose that $\widetilde{\vartheta}_n$ is the maximum likelihood estimator of $\widetilde{\theta}^{(0)}$ based on a random sample $\{X_i\}_{i=1}^n$. Then,

$$\sqrt{n}\left(\widetilde{\vartheta}_n - \widetilde{\theta}^{(0)}\right) \xrightarrow{d} \mathcal{N}\left(0, F^{-1}(\widetilde{\theta}^{(0)})\right)$$

**Estimating Fisher Information**

As before, suppose that $\widetilde{\vartheta}$ is the maximum likelihood estimator of $\widetilde{\theta}^{(0)}$ based on a random sample $\{X_i\}_{i=1}^n$. An estimate of $F(\widetilde{\theta}^{(0)})$ is given by[1]:

$$\begin{aligned}
\widehat{F(\widetilde{\theta}^{(0)})} &= \frac{1}{n}\sum_{i=1}^{n}\left[\left.\frac{\partial \log f(X_i|\widetilde{\theta})}{\partial \widetilde{\theta}}\right|_{\widetilde{\theta}=\widetilde{\vartheta}}\right]\left[\left.\frac{\partial \log f(X_i|\widetilde{\theta})}{\partial \widetilde{\theta}}\right|_{\widetilde{\theta}=\widetilde{\vartheta}}\right]^T \\
&= \left\{\frac{1}{n}\sum_{i=1}^{n}\left.\left(\frac{\partial \log f(X_i|\widetilde{\theta})}{\partial \theta_k}\frac{\partial \log f(X_i|\widetilde{\theta})}{\partial \theta_m}\right)\right|_{\widetilde{\theta}=\widetilde{\vartheta}}\right\}_{k,m=1}^{p}
\end{aligned}$$

[1] *Obviously,* $n\widehat{F(\widetilde{\theta}^{(0)})} = \widehat{I(\widetilde{\theta}^{(0)})}$