2.3.1 Fisher Information

Fisher information is designed to provide a measure of how much information a data set provides about a parameter in a parametric family with some smoothness properties.

Definition 2.78. Suppose that Θ is k-dimensional and that $f_{X|\Theta}(x|\theta)$ is the density of X with respect to ν. The following conditions will be known as the *FI regularity conditions:*

1. *There exists B with $\nu(B) = 0$ such that for all θ, $\partial f_{X|\Theta}(x|\theta)/\partial\theta_i$ exists for $x \notin B$ and each i.*
2. *$\int f_{X|\Theta}(x|\theta)d\nu(x)$ can be differentiated under the integral sign with respect to each coordinate of θ.*
3. *The set $C = \{x : f_{X|\Theta}(x|\theta) > 0\}$ is the same for all θ.*

Definition 2.79. Assume that the three FI regularity conditions above hold. Then the matrix $\mathcal{I}_X(\theta) = ((I_{X,i,j}(\theta)))$ with elements

$$I_{X,i,j}(\theta) = \mathrm{Cov}_\theta\left(\frac{\partial}{\partial\theta_i}\log f_{X|\Theta}(X|\theta), \frac{\partial}{\partial\theta_j}\log f_{X|\Theta}(X|\theta)\right)$$

is called *the Fisher information matrix about Θ based on X.* The random vector with coordinates $\partial\log f_{X|\Theta}(X|\theta)/\partial\theta_i$ is called the *score function.* If T is a statistic, the *conditional score function* is the vector whose ith coordinate is $\partial\log f_{X|T,\Theta}(X|t,\theta)/\partial\theta_i$. The *conditional Fisher information* given $T = t$, denoted by $\mathcal{I}_{X|T}(\theta|t)$, is the conditional covariance matrix of the conditional score function.

Here are some examples.

Example 2.80. Let b be known, and suppose that $X \sim N(\theta, b)$ given $\Theta = \theta$. Then, the FI regularity conditions are satisfied and

$$\begin{aligned} f_{X|\Theta}(x|\theta) &= \frac{1}{\sqrt{2\pi b}}\exp\left\{-\frac{1}{2b}(x-\theta)^2\right\}, \\ \frac{\partial}{\partial\theta}\log f_{X|\Theta}(x|\theta) &= \frac{x-\theta}{b}, \end{aligned}$$

and $\mathcal{I}_X(\theta) = 1/b$. Here we see that the smaller the known variance is, the more information there is in the data about Θ. This is intuitively sensible.

Example 2.81.* Suppose that $X \sim U(0,\theta)$ given $\Theta = \theta$. That is, $f_{X|\Theta}(x|\theta) = \theta^{-1}I_{(0,\theta)}(x)$. In this case FI regularity conditions 1 and 3 fail, but we can still calculate $\partial\log f_{X|\Theta}(x|\theta)/\partial\theta = -1/\theta$. We could then try to define the Fisher information to be the mean of the derivative of this, namely $\mathcal{I}_X(\theta) = 1/\theta^2$. But this function will not have the properties that Fisher information has under all three FI regularity conditions.

*Example 2.81 should actually appear in the text after Example 2.85 on page 113.