**Assumptions 6.2.1** (Additional Regularity Conditions). *Regularity conditions (R3) and (R4) are given by*

**(R3)** *The pdf $f(x;\theta)$ is twice differentiable as a function of $\theta$.*

**(R4)** *The integral $\int f(x;\theta)\,dx$ can be differentiated twice under the integral sign as a function of $\theta$.*

Note that conditions (R1)–(R4) mean that the parameter $\theta$ does not appear in the endpoints of the interval in which $f(x;\theta) > 0$ and that we can interchange integration and differentiation with respect to $\theta$. Our derivation is for the continuous case, but the discrete case can be handled in a similar manner. We begin with the identity

$$1 = \int_{-\infty}^{\infty} f(x;\theta)\,dx.$$

Taking the derivative with respect to $\theta$ results in

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial\theta}\,dx.$$

The latter expression can be rewritten as

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)/\partial\theta}{f(x;\theta)} f(x;\theta)\,dx,$$

or, equivalently,

$$0 = \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,dx. \tag{6.2.1}$$

Writing this last equation as an expectation, we have established

$$E\left[\frac{\partial \log f(X;\theta)}{\partial\theta}\right] = 0; \tag{6.2.2}$$

that is, the mean of the random variable $\frac{\partial \log f(X;\theta)}{\partial\theta}$ is 0. If we differentiate (6.2.1) again, it follows that

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)\,dx + \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,dx. \tag{6.2.3}$$

The second term of the right side of this equation can be written as an expectation, which we call **Fisher information** and we denote it by $I(\theta)$; that is,

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,dx = E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right]. \tag{6.2.4}$$

From equation (6.2.3), we see that $I(\theta)$ can be computed from

$$I(\theta) = -\int_{-\infty}^{\infty} \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)\,dx = -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]. \tag{6.2.5}$$



Using equation (6.2.2), Fisher information is the variance of the random variable $\frac{\partial \log f(X;\theta)}{\partial\theta}$; i.e.,

$$I(\theta) = \operatorname{Var}\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right). \tag{6.2.6}$$

Usually, expression (6.2.5) is easier to compute than expression (6.2.4).

**Remark 6.2.1.** Note that the information is the weighted mean of either

$$\left[\frac{\partial \log f(x;\theta)}{\partial\theta}\right]^2 \quad\text{or}\quad -\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2},$$

where the weights are given by the pdf $f(x;\theta)$. That is, the greater these derivatives are on the average, the more information that we get about $\theta$. Clearly, if they were equal to zero [so that $\theta$ would not be in $\log f(x;\theta)$], there would be zero information about $\theta$. The important function

$$\frac{\partial \log f(x;\theta)}{\partial\theta}$$

is called the **score function**. Recall that it determines the estimating equations for the mle; that is, the mle $\hat\theta$ solves

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial\theta} = 0$$

for $\theta$. ■

**Example 6.2.1** (Information for a Bernoulli Random Variable). Let $X$ be Bernoulli $b(1,\theta)$. Thus

$$\begin{aligned}
\log f(x;\theta) &= x\log\theta + (1-x)\log(1-\theta)\\
\frac{\partial \log f(x;\theta)}{\partial\theta} &= \frac{x}{\theta} - \frac{1-x}{1-\theta}\\
\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
I(\theta) &= -E\left[\frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right]\\
&= \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1}{\theta(1-\theta)},
\end{aligned}$$

which is larger for $\theta$ values close to zero or one. ■

**Example 6.2.2** (Information for a Location Family). Consider a random sample $X_1,\ldots,X_n$ such that

$$X_i = \theta + e_i, \quad i = 1,\ldots,n, \tag{6.2.7}$$