background information on vectors and matrices is given in Appendix B. Additional notation used includes the following. The score vector is the (column) vector $\mathbf{U}(\boldsymbol{\theta}|\mathbf{X}_n)$ with components

$$U_j(\boldsymbol{\theta}|\mathbf{X}_n) = \frac{\partial l(\boldsymbol{\theta}|\mathbf{X}_n)}{\partial \theta_j},$$

and the expected information matrix $\mathbf{I}_n(\boldsymbol{\theta})$ is the matrix whose i, jth component is

$$\mathrm{Cov}[U_i(\boldsymbol{\theta}|\mathbf{X}_n), U_j(\boldsymbol{\theta}|\mathbf{X}_n)].$$

The limiting average information is $\boldsymbol{I}(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\theta})$. As in the 1-dimensional case this will be the information in one observation if the data are iid. The observed information matrix $-\mathbf{H}_n(\boldsymbol{\theta})$ is the matrix whose i, jth component is

$$-\frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{X}_n)}{\partial \theta_i \partial \theta_j}.$$

(This rather odd notation is not standard, but is based on the fact that the matrix of second derivatives of a function of several variables is often referred to as the "Hessian matrix".)

Results on asymptotic theory of MLEs will require a number of "regularity conditions". Various conditions in the following list will be used for different results. They are summarized together here for convenience. Formal conditions are only stated for iid data from a 1-parameter family, although extensions will be briefly discussed in the following sections.

Suppose $X_1, X_2, \ldots$, are iid from a family of pmfs or pdfs (one or the other, but not both)

$$\mathcal{P}_\theta = \{f(x|\theta) : \theta \in \Theta \subseteq R^1\}.$$

The conditions we will need are defined formally as follows:

C.1. θ is identifiable.

C.2. The set $\{x : f(x|\theta) > 0\}$ is the same for all θ.

C.3. The parameter space Θ is an open set.

C.4. $\mathrm{E}_{\theta_0}\{\log[f(X_i|\theta)/f(X_i|\theta_0)]\}$ exists for all $\theta, \theta_0 \in \Theta$.

C.5. With probability 1, $f(x|\theta)$ and $\partial f(x|\theta)/\partial\theta$ are continuous in θ.

C.6. With probability 1, $\partial^2 f(x|\theta)/\partial\theta^2$ and $\partial^3 \log[f(x|\theta)]/\partial\theta^3$ exist and are continuous in θ, and for each $\theta_0 \in \Theta$ there is an open interval B_{θ_0} containing θ_0 and a function $M(x)$ (which can depend on θ_0 but not θ) with $\mathrm{E}_{\theta_0}[M(X_i)] < \infty$, such that $|\partial^3 \log[f(x|\theta)]/\partial\theta^3| \leq M(x)$ for $\theta \in B_{\theta_0}$ (that is, the third derivative of the log likelihood is locally bounded in θ by a function with finite expectation).

C.7.

$$\mathrm{E}_\theta\{\partial \log[f(X_i|\theta)]/\partial\theta\} = 0$$

for all θ.

C.8. The information in one observation $i(\theta) = \mathrm{Var}_\theta\{\partial \log[f(X_i|\theta)]/\partial\theta\}$ exists and $0 < i(\theta) < \infty$ for all θ.

C.9. The expected information can be calculated by taking the expected value of minus the second derivative of the log likelihood, that is

$$\mathrm{Var}_\theta\{\partial \log[f(X_i|\theta)]/\partial\theta\} = \mathrm{E}_\theta\left[-\frac{\partial^2 \log[f(X_i|\theta)]}{\partial\theta^2}\right]$$

for all θ.

Condition C.5 guarantees that the score equation exists and is continuous in θ, and the second derivative portion of C.6 guarantees that the observed information can be computed. Condition C.7 guarantees that the score has mean 0, and C.8 and C.9 guarantee that the information exists and can be computed as minus the expected value of the derivative of the score. Conditions C.7 and C.9 might sound like assuming what you need to prove. These could be replaced by uniform integrability conditions on the derivatives of the density, which would guarantee that the interchange in the order of differentiation and integration used in the derivation of these results can be performed. But in specific applications it is often easier to check directly that C.7 and C.9 are satisfied than it is to verify the uniform integrability conditions, so we leave these as an assumptions to be verified in specific applications, rather than replacing them with more abstract conditions.

6.2 Consistency of MLEs

When an explicit formula is available for the MLE, consistency is usually straightforward to show, as in the following example.

Example 6.2 Consider again the MLE for β in the Weibull distribution from Example 6.1. There it was shown that $\hat{\beta}_n = \sum_{i=1}^n X_i^2/n$. Since the X_i are iid, and