Chapter 6

Maximum Likelihood Methods

6.1 Maximum Likelihood Estimation

Recall in Chapter 4 that as a point estimation procedure, we introduced maximum likelihood estimates (mle). In this chapter, we continue this development showing that these likelihood procedures give rise to a formal theory of statistical inference (confidence and testing procedures). Under certain conditions (regularity conditions), these procedures are asymptotically optimal.

As in Section 4.1, consider a random variable X whose pdf $f(x;\theta)$ depends on an unknown parameter θ which is in a set Ω. Our general discussion is for the continuous case, but the results extend to the discrete case also. For information, we have a random sample (iid) $X_1, \ldots, X_n$ on X. Suppose that $X_1, \ldots, X_n$ are iid random variables with common pdf $f(x;\theta), \theta \in \Omega$. For now, we assume that θ is a scalar, but we do extend the results to vectors in Sections 6.4 and 6.5. The parameter θ is unknown. The basis of our inferential procedures is the likelihood function given by

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Omega, \tag{6.1.1}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$. Because we treat L as a function of θ in this chapter, we have transposed the x_i and θ in the argument of the likelihood function. In fact, we often write it as $L(\theta)$. Actually, the log of this function is usually more convenient to use and we denote it by

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta), \quad \theta \in \Omega. \tag{6.1.2}$$

Note that there is no loss of information in using $l(\theta)$ because the log is a one-to-one function. Most of our discussion in this chapter remains the same if X is a random