Using this, the first partial derivative is

$$l'(\theta) = n - 2\sum_{i=1}^{n} \frac{\exp\{-(x_i-\theta)\}}{1+\exp\{-(x_i-\theta)\}}. \tag{6.1.9}$$

Setting this equation to 0 and rearranging terms results in the equation

$$\sum_{i=1}^{n} \frac{\exp\{-(x_i-\theta)\}}{1+\exp\{-(x_i-\theta)\}} = \frac{n}{2}. \tag{6.1.10}$$

Although this does not simplify, we can show that equation (6.1.10) has a unique solution. The derivative of the left side of equation (6.1.10) simplifies to

$$(\partial/\partial\theta)\sum_{i=1}^{n} \frac{\exp\{-(x_i-\theta)\}}{1+\exp\{-(x_i-\theta)\}} = \sum_{i=1}^{n} \frac{\exp\{-(x_i-\theta)\}}{(1+\exp\{-(x_i-\theta)\})^2} > 0.$$

Thus the left side of equation (6.1.10) is a strictly increasing function of θ. Finally, the left side of (6.1.10) approaches 0 as $\theta \to -\infty$ and approaches n as $\theta \to \infty$. Thus equation (6.1.10) has a unique solution. Also, the second derivative of $l(\theta)$ is strictly negative for all θ; so the solution is a maximum.

Having shown that the mle exists and is unique, we can use a numerical method to obtain the solution. In this case, Newton's procedure is useful. We discuss this in general in the next section, at which time we reconsider this example. ■

Example 6.1.3. In Example 4.1.2, we discussed the mle of the probability of success θ for a random sample $X_1, X_2, \ldots, X_n$ from the Bernoulli distribution with pmf

$$p(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & x = 0, 1 \\ 0 & \text{elsewhere,} \end{cases}$$

where $0 \le \theta \le 1$. Recall that the mle is $\overline{X}$, the proportion of sample successes. Now suppose that we know in advance that, instead of $0 \le \theta \le 1$, θ is restricted by the inequalities $0 \le \theta \le 1/3$. If the observations were such that $\overline{x} > 1/3$, then $\overline{x}$ would not be a satisfactory estimate. Since $\frac{\partial l(\theta)}{\partial \theta} > 0$, provided $\theta < \overline{x}$, under the restriction $0 \le \theta \le 1/3$, we can maximize $l(\theta)$ by taking $\widehat{\theta} = \min\{\overline{x}, \frac{1}{3}\}$. ■

The following is an appealing property of maximum likelihood estimates.

Theorem 6.1.2. *Let $X_1, \ldots, X_n$ be iid with the pdf $f(x;\theta), \theta \in \Omega$. For a specified function g, let $\eta = g(\theta)$ be a parameter of interest. Suppose $\widehat{\theta}$ is the mle of θ. Then $g(\widehat{\theta})$ is the mle of $\eta = g(\theta)$.*

Proof: First suppose g is a one-to-one function. The likelihood of interest is $L(g(\theta))$, but because g is one-to-one,

$$\max L(g(\theta)) = \max_{\eta = g(\theta)} L(\eta) = \max_{\eta} L(g^{-1}(\eta)).$$

But the maximum occurs when $g^{-1}(\eta) = \widehat{\theta}$; i.e., take $\widehat{\eta} = g(\widehat{\theta})$.

Suppose g is not one-to-one. For each η in the range of g, define the set (preimage)

$$g^{-1}(\eta) = \{\theta : g(\theta) = \eta\}.$$

The maximum occurs at $\widehat{\theta}$ and the domain of g is Ω, which covers $\widehat{\theta}$. Hence, $\widehat{\theta}$ is in one of these preimages and, in fact, it can only be in one preimage. Hence to maximize $L(\eta)$, choose $\widehat{\eta}$ so that $g^{-1}(\widehat{\eta})$ is that unique preimage containing $\widehat{\theta}$. Then $\widehat{\eta} = g(\widehat{\theta})$. ■

Consider Example 4.1.2, where $X_1, \ldots, X_n$ are iid Bernoulli random variables with probability of success p. As shown in this example, $\widehat{p} = \overline{X}$ is the mle of p. Recall that in the large sample confidence interval for p, (4.2.7), an estimate of $\sqrt{p(1-p)}$ is required. By Theorem 6.1.2, the mle of this quantity is $\sqrt{\widehat{p}(1-\widehat{p})}$.

We close this section by showing that maximum likelihood estimators, under regularity conditions, are consistent estimators. Recall that $\mathbf{X}' = (X_1, \ldots, X_n)$.

Theorem 6.1.3. *Assume that $X_1, \ldots, X_n$ satisfy the regularity conditions (R0) through (R2), where θ_0 is the true parameter, and further that $f(x;\theta)$ is differentiable with respect to θ in Ω. Then the likelihood equation,*

$$\frac{\partial}{\partial\theta} L(\theta) = 0,$$

or equivalently

$$\frac{\partial}{\partial\theta} l(\theta) = 0,$$

has a solution $\widehat{\theta}_n$ such that $\widehat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: Because θ_0 is an interior point in Ω, $(\theta_0 - a, \theta_0 + a) \subset \Omega$, for some $a > 0$. Define S_n to be the event

$$S_n = \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 - a; \mathbf{X})\} \cap \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 + a; \mathbf{X})\}.$$

By Theorem 6.1.1, $P(S_n) \to 1$. So we can restrict attention to the event S_n. But on S_n, $l(\theta)$ has a local maximum, say, $\widehat{\theta}_n$, such that $\theta_0 - a < \widehat{\theta}_n < \theta_0 + a$ and $l'(\widehat{\theta}_n) = 0$. That is,

$$S_n \subset \left\{\mathbf{X} : |\widehat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\widehat{\theta}_n(\mathbf{X})) = 0\right\}.$$

Therefore,

$$1 = \lim_{n\to\infty} P(S_n) \le \overline{\lim_{n\to\infty}}\, P\left[\left\{\mathbf{X} : |\widehat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\widehat{\theta}_n(\mathbf{X})) = 0\right\}\right] \le 1;$$

see Remark 5.2.3 for discussion on $\overline{\lim}$. It follows that for the sequence of solutions $\widehat{\theta}_n$, $P[|\widehat{\theta}_n - \theta_0| < a] \to 1$.

The only contentious point in the proof is that the sequence of solutions might depend on a. But we can always choose a solution "closest" to θ_0 in the following