LECTURE 1

Some Examples of Maximum Likelihood Estimation

Maximum Likelihood Estimation is a statistical estimation technique that chooses as estimates of the parameters of a statistical model the parameter values that, given the data, maximize the "likelihood" that the observed data was generated from a given model.

Example 1: Consider a binomial experiment of n trials resulting in the observations $x_1, x_2, ..., x_n$, where $x_i = 1$ if the trial was a success, $x_i = 0$ otherwise. The likelihood of the observed sample $x_1, x_2, ..., x_n$ is the product of the likelihoods of the individual observations x_i, namely,

$$L(x_1, x_2, \cdots, x_n; p) = L(x_1, p) \cdot L(x_2 : p) \cdots L(x_n; p)$$

$$= p^{x_1} \cdot (1-p)^{1-x_1} \cdot p^{x_2} \cdot (1-p)^{1-x_2} \cdot ... \cdot p^{x_n} \cdot (1-p)^{1-x_n}$$

$$= p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$= p^{X} \cdot (1-p)^{n-X} \qquad (1)$$

where $X = \sum_{i=1}^{n} x_i$ and p is the unknown probability of success. We have assumed the independence of the observations $x_1, x_2, ..., x_n$ and hence the ability to factor the likelihood of the sample $L(x_1, x_2, ..., x_n; p)$ into its respective parts $L(x_1; p), L(x_2; p), ..., L(x_n; p)$.

We note that ln(L) is a monotonically increasing function of the likelihood L and thus both L and ln(L) are maximized at the same value of p.

Since it is tedious to take the derivatives of products of terms, let's work on maximizing the log likelihood function.

$$\ell = \ln(L) = X \ln(p) + (n - X)\ln(1 - p) \qquad (2)$$

(Note: we restrict $0<p<1$, otherwise (2) is not defined)

Taking the first derivative of ln(L) with respect to p we get

$$\frac{d\ln(L)}{dp} = X \cdot \frac{1}{p} - (n-X) \cdot \frac{1}{1-p}$$
$$= \frac{X}{p} - \frac{(n-X)}{1-p} \quad . \qquad (3)$$

The first order condition is

$$\frac{X}{\hat{p}} - \frac{(n-X)}{1-\hat{p}} = 0 \qquad (4)$$

where $\hat{p}$ denotes that value of p which maximizes the log likelihood function (2), and hence maximizes the likelihood function (1).

The solution to the first order condition (4) is:

$$\hat{p} = \frac{X}{n} \qquad (5)$$

Now to see if $\hat{p}$ is associated with a maximum of (2) or (1) we need to examine the second order condition. The second derivative of the log likelihood function is

$$\frac{d^2\ln(L)}{dp^2} = \frac{d}{dp}\left[\frac{X}{p} - \frac{(n-X)}{1-p}\right]$$
$$= -\frac{X}{p^2} - \frac{(n-X)}{(1-p)^2} \qquad (6)$$

(Recall 0<p<1 is assumed.) Therefore, from (6) we can see than $\frac{d^2\ln(L)}{dp^2} < 0$ and the second order condition for a maximum is satisfied.

Example 2: Let $x_1, x_2, ..., x_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Find the maximum likelihood estimators of μ and σ^2. The likelihood function is:

$$L(x_1, x_2, ..., x_n; \mu, \sigma^2) = f(x_1, x_2, ..., x_n) = f(x_1) \cdot f(x_2) \cdot ... \cdot f(x_n)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_1-\mu)^2}{2\sigma^2}\right\} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_2-\mu)^2}{2\sigma^2}\right\}$$

$$\cdot \ldots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_n-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\} \qquad (7)$$

And the log likelihood function is:

$$\ell = \ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \qquad (8)$$

Taking derivatives of the log likelihood function with respect to μ and σ^2 yields

$$\frac{\partial\ell}{\partial\mu} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} \qquad (9)$$

and

$$\frac{\partial\ell}{\partial\sigma^2} = -\left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{\sum(x_i-\mu)^2}{2\sigma^4} \ . \qquad (10)$$

Setting the derivatives (9) and (10) to zero and solving for the unknown $\hat{\mu}$ and $\hat{\sigma}^2$ will yield the maximum likelihood estimates for μ and σ^2. Let's begin with the first, first-order condition.

$$\frac{\sum_{i=1}^{n}(x_i-\hat{\mu})}{\hat{\sigma}^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\hat{\mu}) = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i - n\hat{\mu} = 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n}x_i}{n} = \bar{x} \ . \qquad (11)$$

Thus, the sample mean $\bar{x}$ is the maximum likelihood estimator of μ. Now let us turn to the maximum likelihood estimator of σ^2.

The second, first-order condition becomes

$$\frac{\partial \ell}{\partial \sigma^2} = -\left(\frac{n}{2}\right)\left(\frac{1}{\hat{\sigma}^2}\right) + \frac{\sum (x_i - \hat{\mu})^2}{2\hat{\sigma}^4} = 0 \qquad (12)$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \quad . \qquad (13)$$

Thus, the maximum likelihood estimator of σ^2 is (13) and is biased, but is consistent.

To verify that (11) and (13) are stationary points that provide a maximum as opposed to a minimum or a saddle point, we need to examine the Hessian matrix of the system. The Hessian is

$$H = \begin{bmatrix} \dfrac{\partial^2 \ell}{\partial \mu^2} & \dfrac{\partial^2 \ell}{\partial \mu \partial \sigma^2} \\ \dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \mu} & \dfrac{\partial^2 \ell}{(\partial \sigma^2)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{-n}{\sigma^2} & \dfrac{-\sum_{i=1}^{n}(x_i - \mu)}{\sigma^4} \\ \bullet & \dfrac{n}{2\sigma^4} - \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^6} \end{bmatrix} \quad . \qquad (14)$$

Let H($\hat{\mu}, \hat{\sigma}^2$) denote the Hessian matrix evaluated at the maximum likelihood estimates and $H_{ij}(\hat{\mu}, \hat{\sigma}^2)$ be the *i,j*-th element of that matrix. Sufficient conditions for $\hat{\mu}$ and $\hat{\sigma}^2$ to provide a local maximum of the log likelihood function (8) are

(i) $(\hat{\mu}, \hat{\sigma}^2)$ provides a stationary point (as was shown in the previously derived first order conditions)

(ii) $H_{11}(\hat{\mu}, \hat{\sigma}^2) < 0$ and D=det$[H(\hat{\mu}, \hat{\sigma}^2)] > 0$.

Note condition (ii) simply states that the principal minors of the Hessian matrix evaluated at the stationary point must be alternating in sign, with the first principal minor being negative. The conditions in (ii) are satisfied in that $H_{11}(\hat{\mu}, \hat{\sigma}^2) = -\frac{n}{\hat{\sigma}^2} < 0$ and D = $\frac{n^2}{2\hat{\sigma}^6} > 0$.