LECTURE - 19

Agenda:

1. Method of maximum likelihood
2. Examples

We have seen two methods to compute estimators. The method of sufficient statistics and the method of moments. Both of these methods have drawbacks. In the method of sufficient statistics, we need to find a function of the sufficient statistic which is unbiased, which can be difficult. The method of moments is intuitive and easy to use, but does not always give the "best" estimators. We will now present the method of maximum likelihood, which often leads to the "best" estimators.

The principle behind this method is very simple. "CHOOSE THE PARAMETER ESTIMATE, THAT MAXIMIZES THE PROBABILITY OF OBTAINING THE OBSERVED SAMPLE."

Let us look at a simple example before proceeding to the general method.

Suppose we have a box with 3 balls. We know that each of the balls may be red or white, but we do not know the total number of either color.

<u>TASK:</u> Estimate the number of red balls in the box.

<u>Data:</u> Two balls are randomly chosen without replacement from the box. Both the balls are red.

$$P(\text{Two red balls in sample} \mid \text{Total red balls} = 0) = 0.$$

$$P(\text{Two red balls in sample} \mid \text{Total red balls} = 1) = 0.$$

$$P(\text{Two red balls in sample} \mid \text{Total red balls} = 2) =$$

$$\frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = \frac{1}{3}.$$

$$P(\text{Two red balls in sample} \mid \text{Total red balls} = 3) = \frac{\binom{3}{2}}{\binom{3}{2}} = 1.$$

Hence, the probability of the observed data is maximized if the total number of red balls is 3. Hence, the maximum likelihood estimate of the total number of red balls in the box is 3.

GENERAL PROCEDURE:

- Suppose we have a sample $Y_1, Y_2, ---, Y_n$ from a population with parameter θ. Suppose the specific configuration that has been observed in a particular instance is $y_1, y_2, ---, y_n$.

① Write down the likelihood $L(y_1, y_2, ---, y_n | \theta)$.

② Choose $\hat{\theta}$ as the value of θ which maximizes $L(y_1, y_2, ---, y_n | \theta)$, i.e.,

$$\hat{\theta} = \arg\min_{\theta} L(y_1, y_2, ---, y_n | \theta).$$

Example 1: Suppose n independent tosses of a coin resulted in observations $y_1, y_2, ---, y_n$ in a particular instance.
Find the maximum likelihood estimate of p, the underlying probability of heads.

Recall that for discrete random variables, the likelihood of the observed configuration is the joint probability mass function at the observed configuration. Hence,

$$L(y_1, y_2, \ldots, y_n \mid p) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid p)$$
$$= p^{\sum_{i=1}^{n} y_i} (1-p)^{n - \sum_{i=1}^{n} y_i}.$$

For the sake of simplicity of notation, let us denote $L(y_1, y_2, \ldots, y_n \mid p)$ by $L(p)$ and let us denote $\sum_{i=1}^{n} y_i$ by y. Hence,

$$L(p) = p^{y} (1-p)^{n-y}.$$

WE NOW WISH TO FIND THE VALUE OF p THAT MAXIMIZES $L(p)$.

<u>CASE I</u>: If $y = 0$, i.e., all tosses are tails, then $L(p) = (1-p)^n$ is maximized at $p = 0$. Hence $\boxed{\hat{p} = 0 = \frac{y}{n}}$.

<u>CASE II</u>: If $y = n$, i.e., all tosses are heads, then $L(p) = p^n$ is maximized at $p = 1$. Hence $\boxed{\hat{p} = 1 = \frac{y}{n}}$.

CASE III: $y = 1, 2, ---, n-1$.

We use the calculus principle that the value of p that maximizes $L(p)$ can be found by setting the derivative $\frac{dL(p)}{dp}$ to 0 and then solving for p.

$$\frac{dL(p)}{dp} = y p^{y-1} (1-p)^{n-y} - (n-y) p^{y} (1-p)^{n-y-1}.$$

Hence,

$$\frac{dL(p)}{dp} = 0$$

$$\Leftrightarrow \quad y p^{y-1}(1-p)^{n-y} - (n-y) p^{y} (1-p)^{n-y-1} = 0$$

$$\Leftrightarrow \quad \left\{ y(1-p) - (n-y)p \right\} p^{y-1} (1-p)^{n-y-1} = 0$$

$$\Leftrightarrow \quad y(1-p) - (n-y)p = 0$$

$$\text{or} \quad p = 0$$

$$\text{or} \quad 1 - p = 0$$

$$\Leftrightarrow \quad p = \frac{y}{n} \quad \text{or} \quad p = 0 \quad \text{or} \quad p = 1.$$

Note that $L(0) = 0$ and $L(1) = 0$ in this case. Hence $p = 0$ and $p = 1$ are not candidates for the maximum of $L(p)$.

One can easily verify that $\frac{d^2}{dp^2} L\left(\frac{y}{n}\right) < 0$,

and hence $\boxed{\hat{p} = \frac{y}{n}}$ maximizes $L(p)$.

Hence, the maximum likelihood estimate of p for an observed configuration $y_1, y_2, \ldots, y_n$ is

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} y_i = \text{Observed Sample proportion}.$$