Maximum Likelihood Estimation

We've used $\bar{x}$ to estimate μ and $\hat{p}$ to estimate p, because it seems "correct". Can we get a better justification?

Yes.

In regression, we discussed parameter estimation using **least squares**.

Another method of estimation is **maximum likelihood estimation**: what value of the parameter(s) will maximize the probability (or likelihood) of obtaining a particular sample?

NOTE: Maximize function = calculus (usually).

Let's consider the following example.

Ex: Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n drawn from a Binomial$(1, p)$ distribution, where p is unknown. What is the maximum likelihood estimate (**MLE**) of p?

Solution: The probability mass function of X is

$$\begin{aligned} P(X = x) &= \binom{1}{x} p^x (1-p)^{1-x}, \quad x = 0, 1 \\ &= p^x (1-p)^{1-x}. \end{aligned}$$

Since we have a random sample, the $X_i's$ are independent. Therefore we can write

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) &= P(X_1 = x_1) P(X_2 = x_2) \ldots P(X_n = x_n) \\ &= p^{x_1}(1-p)^{1-x_1} \ldots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} \\ &= L(p) \end{aligned}$$

where $L(p)$ is the **likelihood function**.

We want to find p to maximize $L(p)$: this is **hard**.

Easier: Work with $\ln[L(p)] = l(p)$, the **log likelihood**.

$$\begin{aligned} l(p) &= \ln[p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}] \\ &= \ln[p^{\sum_i x_i}] + \ln[(1-p)^{n-\sum_i x_i}] \\ &= (\sum_i x_i) \ln(p) + (n - \sum_i x_i) \ln(1-p) \end{aligned}$$

To maximize $l(p)$ with respect to p, do the usual calculus stuff: take derivative of $l(p)$ with respect to p, set result equal to 0, and solve for p.

$$
\begin{aligned}
\frac{dl}{dp} &= \sum_i x_i \left(\frac{1}{p}\right) + (n - \sum_i x_i)\left(\frac{-1}{1-p}\right) = 0 \\
\frac{\sum_i x_i}{p} &= \frac{(n - \sum_i x_i)}{1-p} \\
(1-p)(\sum_i x_i) &= p(n - \sum_i x_i) \\
p(n - \sum_i x_i + \sum_i x_i) &= \sum_i x_i \\
\hat{p} &= \frac{\sum_i x_i}{n}. \quad \square
\end{aligned}
$$

Now, recall that $x_i = 0$ or $x_i = 1$. Therefore, $\sum_i x_i$ is the total number of "successes", or the number of outcomes of interest.

Therefore, $\hat{p}$ is the x/n that we used before.

Text, pg. 588-589: MLE's have many desirable properties.