Maximum Likelihood Estimation

Maximum Likelihood Criterion

- We have seen different criteria for goodness of a point estimator.
- Now, we will see another criterion which is based on the joint distribution of the random variables whose realizations are the data.

Suppose we have a parameter θ.

Case 1. Discrete random variables:

- Under different possible values of the parameter θ, we can calculate the probability of getting the values of the random variables that we observe.

- If we compare all these probabilities, then the value of θ corresponding to the maximum probability is most likely among all values of θ.
- This value of θ is the maximum likelihood estimate of θ for the the given data.
- Note that this estimate of θ depends on the data. In other words, this estimate is a function of the data (observed random variables).
- So, the maximum likelihood estimator (MLE) is a statistic.

Case 2. Continuous random variables:

- In the continuous case the probability of any single value of a random variable is zero.
- So, we compare the probability densities (instead of probabilities) for different values of θ at the observed value of the random variables and take the value of θ for which the density is maximum.

Likelihood Function

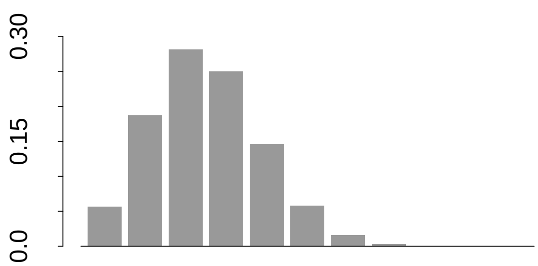
- The likelihood function $L(y \mid \theta)$ is the joint p.m.f. of the data when the random variables are discrete and the joint p.d.f. of data when the random variables are continuous.
- The likelihood function is viewed as a function of the parameter θ for a fixed value of the observed random variables.
- So, often we will write $L(\theta)$ for $L(y \mid \theta)$.

Maximum Likelihood Estimator (MLE)

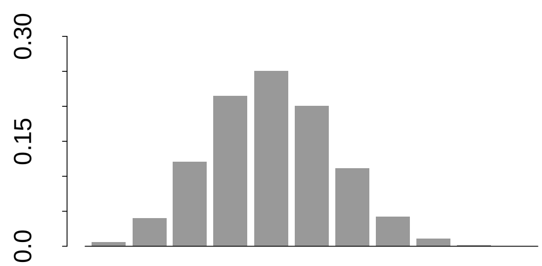
Formally, the maximum likelihood estimator is the value of θ, expressed as a statistic (function of the observed random variables y), that maximizes the likelihood function $L(y \mid \theta)$.

Binomial example We have 10 people receiving a new treatment for a serious disease. What is the probability of recovery (success) p?

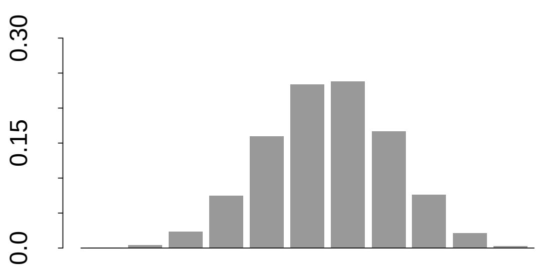
Binomial Probability Histograms



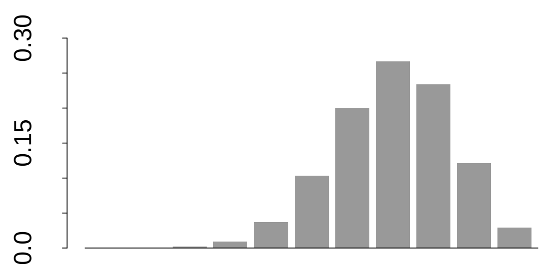
Binomial prob. hist. with n=10 and p=0.25



Binomial prob. hist. with n=10 and p=0.4



Binomial prob. hist. with n=10 and p=0.55



Binomial prob. hist. with n=10 and p=0.7

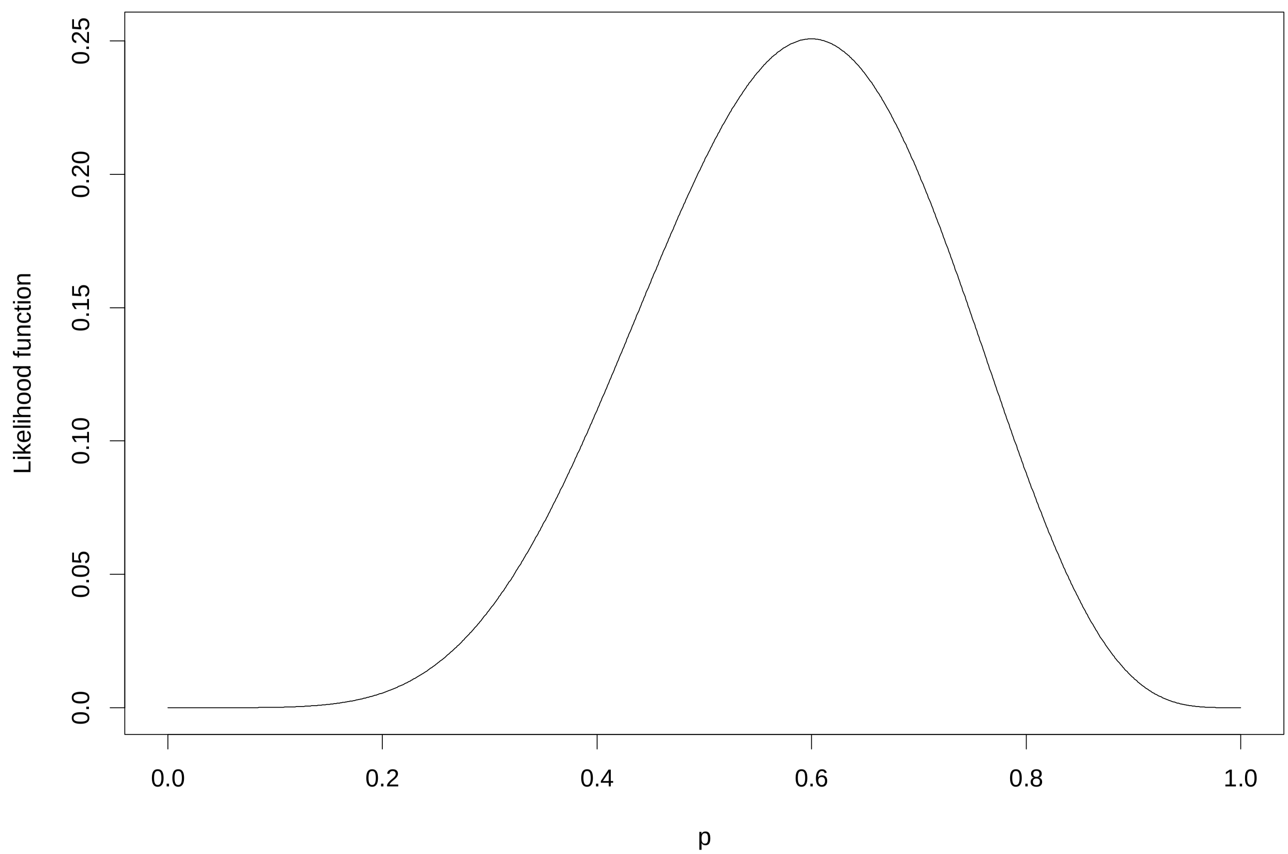
Maximum Likelihood Methodology for the Example

Assume that 6 of the 10 patients receiving the new treatment recover.

- Write down the probability distribution for the number of successes in n trials. (This is the likelihood function $L(p)$.)

- Plug the information from your sample in the likelihood function.

Likelihood Function for p



- Find the value of p that maximizes $L(p)$.

- This is your *maximum likelihood estimate* for p.

Maximum Likelihood Methodology

Let's work through the more general case and find a formula for the maximum likelihood estimate for p - a formula which depends just on observed data.

- As before, write down likelihood function.

- How to take the derivative, set it equal to zero, and solve for p?

- Find the value of p that maximizes $L(p)$.

- This is your *maximum likelihood estimate* for p.

MLE for Mean of the Normal

Assume that we have a random sample $x_1, x_2, ..., x_n$ from a normally distributed population for which the variance σ^2 is known. What is the MLE for the mean μ? The probability density function for the normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2] \ , \quad -\infty < x < \infty \ , -\infty < \mu < \infty$$

MLE for Variance of the Normal

Assume that we have a random sample $x_1, x_2, ..., x_n$ from a normally distributed population for which the mean μ is known. What is the MLE for the variance σ^2? The probability density function for the normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2] \ , \quad -\infty < x < \infty \ , \sigma^2 > 0$$

MLE for Uniform (0,θ)

Assume that we have a random sample $x_1, x_2, ..., x_n$ from a population which is distributed uniformly on [0,θ]. What is the MLE for θ?

Invariance Property of MLEs

- Suppose we know the MLE of a parameter and want to find the MLE for another parameter which is a function of the former parameter.
- The MLE for the latter parameter can be found by plugging the MLE for the former parameter into the function that gives the latter parameter.
- If $t(\theta)$ is a function of θ and $\hat{\theta}$ is the MLE for θ, then the MLE for $t(\theta)$ is given by $t(\hat{\theta})$.

Example Using the Invariance Property

We know that for a binomial proportion of successes p, the MLE is given by $\hat{p}$. What is the MLE for the variance of Y?

Example

Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a distribution for which the p.d.f. $f(x|\theta)$ is given below. Also, suppose θ is unknown ($\theta > 0$). Find the MLE of θ.

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & \text{for} \quad 0 < x < 1 \\ \\ 0 & \text{otherwise} \end{cases}$$