Definition – Statistical Inference

In statistics, **inference** is the process of using data from a **sample** to draw conclusions or make predictions about a **population**.

There are two main types of statistical inference:

- **Estimation** – estimating population parameters (e.g., proportion) using sample data.
- **Hypothesis testing** – evaluating a claim about population parameters (e.g., proportion) based on sample data.

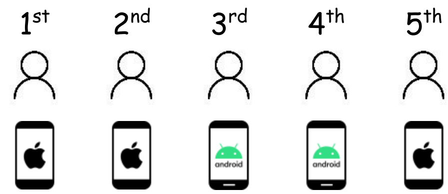
In this lecture, we will be learning how to estimate a population proportion the method of **Maximum Likelihood Estimation**.

Contents

Estimation of a Proportion Using the Method of Maximum Likelihood

Example – Suppose we want to estimate the proportion of iPhone users in a population.
Let $\boldsymbol{p}$ the true proportion of iPhone users in the population (unknown).

To do that, we take a random sample of 5 people and observe 3 of them use iPhone (the other two does not use it). Details are shown below.

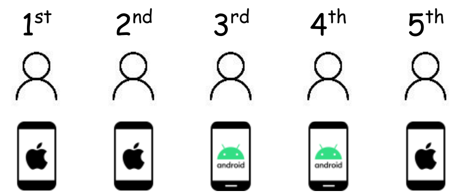


Question: What is a reasonable estimate of the population proportion, $\boldsymbol{p}$?

Key Idea of Maximum Likelihood

The idea of **maximum likelihood estimation (MLE)** is simple: Among all possible values of $\boldsymbol{p}$**,** we choose the value of $\boldsymbol{p}$ that **maximizes** the probability of producing our observed data.

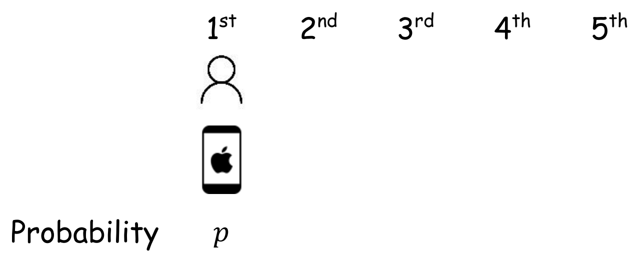
In this example, we want the value of $\boldsymbol{p}$ that gives the **highest probability of producing the data we actually observed (3 use iPhone and the other 2 do not use it)**.



Let's derive the probability of producing our observed data.

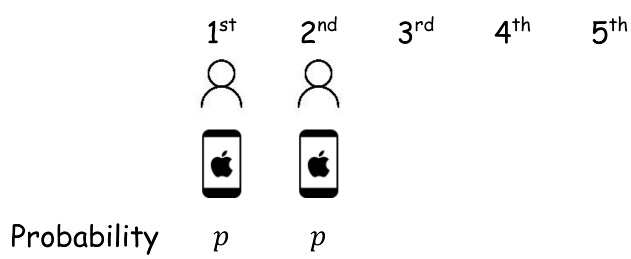
Recall, the proportion of iPhone users in the population is given by p.

So, when we pick the 1st person at random, the probability that this person uses iPhone is given by p



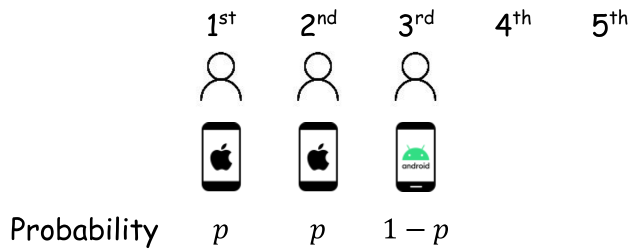
When we pick the 2nd person at random, the probability that this person uses iPhone is also p

Note: *We assume that whether the first person uses an iPhone does not affect whether the second person uses an iPhone (it is called independence condition).*

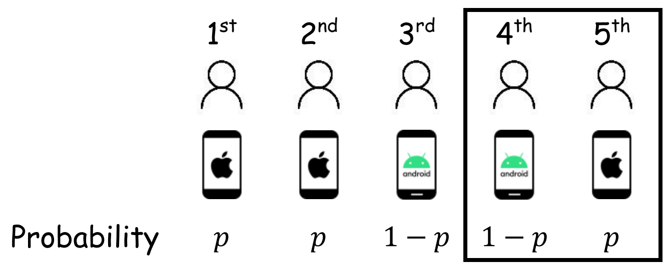


When we pick the 3rd person at random, the probability that this person DOES NOT use iPhone is given by $1 - p$

Note: *We assume that whether the first / second person use an iPhone does not affect whether the second person uses an iPhone (it is called independence condition).*



Following the same logic, the probability that the 4th person does **not** use an iPhone is $1 - p$, and the probability that the 5th person use an iPhone is p.



Therefore, the probability of getting our observed data is given by:

$$P\left(\begin{matrix}1st\\ uses\\ iPhone\end{matrix}\quad \textbf{and}\quad \begin{matrix}2nd\\ uses\\ iPhone\end{matrix}\quad \textbf{and}\quad \begin{matrix}3rd\\ doesn't\ use\\ iPhone\end{matrix}\quad \textbf{and}\quad \begin{matrix}4th\\ doesn't\ use\\ iPhone\end{matrix}\quad \textbf{and}\quad \begin{matrix}5th\\ use\\ iPhone\end{matrix}\right)$$

=

[By independent rule]

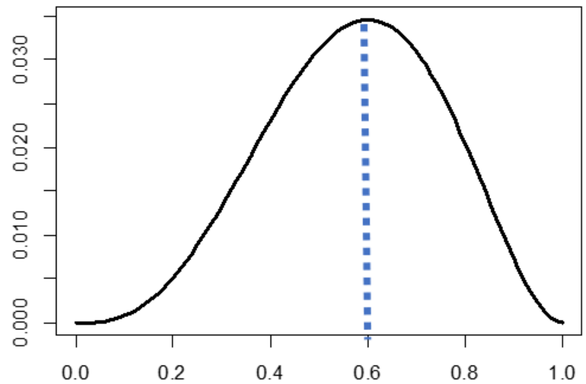
=

=

When we view this probability as a function of the unknown parameter p, it is called the **likelihood function**, denoted by $L(p)$

$$L(p) = p^3(1-p)^2$$

By graphing this function, we can see that $L(p)$ has a maximum at some value of p.



Our goal is to find the value of p that **maximizes** this likelihood function

$$L(p) = p^3(1-p)^2$$

Important:

To make the math easier, we maximize the **log-likelihood** instead of the likelihood itself.

This does not change the location of the maximum.

$$\boldsymbol{\ell}(p) = \ln L(p)$$

$$=$$

Take the derivative with respect to p:

$$\frac{d}{dp}\boldsymbol{\ell}(p) =$$

$$=$$

Set the derivative equal to zero and solve for p.
Let $\bar{p}$ denote the value of p that makes the derivative equal to zero.

$$0 = \boldsymbol{\ell}'(\bar{p})$$

$$0 = \frac{3}{\bar{p}} - \frac{2}{1-\bar{p}}$$

After some algebra, we find out that

$$\bar{p} = \frac{3}{3+2} = \frac{3}{5} = 0.6$$

So, the value of p that maximizes the likelihood (or log-likelihood) function is 0.6

This is exactly **the proportion of iPhone users we observed in the sample**

So, based on our data, we estimate that **about 60% of the population uses an iPhone**.

Question: Suppose we observe a sample of 5 people in which exactly 4 individuals use an iPhone and the other 1 does not. However, we do not observe which specific individuals use an iPhone.
Formulate the **likelihood function** for this observation and find the **maximum likelihood estimate**.

Solution: We need to consider all possible cases in which 4 people use an iPhone and 1 person does not, and compute the probability of each case.

1st 2nd 3rd 4th 5th

Probability

Case 1 

Case 2

Case 3

Case 4

Case 5

So, the probability of producing our data (4 use an iPhone and 1 does not)

=

The likelihood function is given by:

General Results of Maximum Likelihood Estimation for a Proportion

In the population, $\boldsymbol{p} \times \mathbf{100\%}$ of all individuals falls into the category of interest.
(e.g. 60% of all people use an iPhone)

Suppose we take a random sample of n individuals and observe

- x individuals fall into the category of interest (e.g use an iPhone) and
- the remaining $n - x$ individuals do not fall into the category of interest (e.g. do not use an iPhone)

The likelihood function, that the probability of producing our data - categorical data with two categories (e.g. observing $\boldsymbol{x}$ iphone users and $\boldsymbol{n} - \boldsymbol{x}$ non-iPhone users) is given:

$$L(p) = constant \ * \ p^x (1-p)^{n-x}$$

Important: In the function,
- the data (n,x) are fixed
- the parameter (p) **varies**.

The corresponding log-likelihood function is given by:

$$\boldsymbol{\ell}(p) = x \ln p + (n-x)\ln(1-p) + \ln constant$$

The derivative of the log-likelihood function is given by

$$\boldsymbol{\ell}(p) = \frac{x}{p} - \frac{n-x}{1-p}$$

$$= \frac{x - np}{p(1-p)}$$

The value of p that **maximizes** the likelihood function, denoted by $\bar{p}$ is given by:

$$\bar{p} = \frac{x}{n}$$

This is exactly the sample proportion of individuals who fall into the category of interest.

Key Takeaway

- The **maximum likelihood estimator** of a population proportion is the **sample proportion**.
- This explains why we use the sample proportion to **estimate** and **test** a population proportion.

This method is called the **Maximum Likelihood Estimation (MLE)**