



MAXIMUM LIKELIHOOD ESTIMATES OF ALLELES

Dr. R. Sivaraman

Associate Professor, Department of Mathematics, D. G. Vaishnav College, Chennai, India
National Awardee for Popularizing Mathematics among masses

Email: rsivaraman1729@yahoo.co.in

DOI: [10.33329/bomsr.9.1.1](https://doi.org/10.33329/bomsr.9.1.1)



ABSTRACT

In this paper, I have used the concept of logarithm of Maximum Likelihood function to discuss the properties of two particular alleles for a particular genotype. The main purpose of this paper is to derive the frequencies of the two considered alleles. In the final section, I had provided an example to verify the results obtained. I had made use of the assumption given by Hardy – Weinberg Principle for deriving the frequencies of the alleles.

Keywords: Frequency of Alleles, Hardy – Weinberg Principle, Population Frequencies, Logarithm of Likelihood Function, Maximum Likelihood Estimate.

1. Introduction

The concept of likelihood function is used to determine the frequencies of two alleles for a population of a diploid organism in a particular gene. After constructing the likelihood function, we consider its logarithm and then determine the logarithm of the likelihood function which we call log-likelihood to eventually find the frequencies of the alleles considered. These simple calculations will be useful for having ideas about interaction of alleles in population of genes.

2. Setting the Stage

We will consider a population of a diploid organism which has two alleles for a particular gene, denoted by A and a respectively. If the population is in Hardy – Weinberg equilibrium, and the frequency of the alleles are p_A and $p_a = 1 - p_A$, then the genotypes of individuals in the population should have the following frequencies $AA : p_A^2$, $Aa : 2p_A p_a$, $aa : p_a^2$ (2.1)

3. Determination of Allele Frequencies

In this section, I will explain the method of determining Allele frequencies using the information obtained from (2.1). In particular, I will demonstrate the technique for determining maximum Likelihood estimates of the allele frequencies p_A and p_a .

First, let us consider a random sample of n individuals in a population which have two alleles of types A and a . Let n_{AA}, n_{Aa}, n_{aa} be the number of individuals in the population whose frequencies are given by (2.1).

Since there are n individuals in the population whose allele frequencies satisfy (2.1), by Hardy – Weinberg principle, we have

$$np_A^2 = n_{AA} \text{ giving } p_A^2 = \frac{n_{AA}}{n} \quad (3.1)$$

$$n \times 2p_A(1 - p_A) = n_{Aa} \text{ giving } 2p_A(1 - p_A) = \frac{n_{Aa}}{n} \quad (3.2)$$

$$n \times (1 - p_A)^2 = n_{aa} \text{ giving } (1 - p_A)^2 = \frac{n_{aa}}{n} \quad (3.3)$$

We thus have three equations (3.1), (3.2) and (3.3) with one unknown p_A .

To determine the maximum frequencies of Alleles considered, I consider the logarithm of Likelihood function of p_A . By definition of Likelihood function, we have

$$\ln(p_A) = n_{AA} \ln(p_A^2) + n_{Aa} \ln(2p_A(1 - p_A)) + n_{aa} \ln((1 - p_A)^2) \quad (3.4)$$

Using the properties of logarithms, we have

$$\ln(p_A) = 2n_{AA} \ln(p_A) + n_{Aa} [\ln 2 + \ln(p_A) + \ln(1 - p_A)] + 2n_{aa} \ln(1 - p_A) \quad (3.5)$$

Now to find the maximum likelihood estimate of p_A we have to differentiate $\ln(p_A)$ with respect to the parameter p_A and equate it to zero. Doing this, we get

$$\frac{2n_{AA}}{p_A} + \frac{n_{Aa}}{p_A} - \frac{n_{Aa}}{1 - p_A} - \frac{2n_{aa}}{1 - p_A} = 0 \quad (3.6)$$

Solving for p_A from (3.6), we have $(2n_{AA} + n_{Aa})(1 - p_A) = (n_{Aa} + 2n_{aa})p_A$ giving

$$p_A = \frac{2n_{AA} + n_{Aa}}{2(n_{AA} + n_{Aa} + n_{aa})} \quad (3.7)$$

The value of from (3.7) is the maximum likelihood estimate provides the frequency of the first allele among two alleles that we have considered. The frequency for the second allele is given by

$$p_a = 1 - p_A = \frac{2n_{aa} + n_{Aa}}{2(n_{AA} + n_{Aa} + n_{aa})} \quad (3.8)$$

Expressions (3.7) and (3.8) provide the frequencies which maximize the likelihood function for the two alleles considered.

4. Verification

In this section, we verify the results obtained by considering $n_{AA} = 2n_{Aa} = n_{aa} = 1$. In this case, the logarithm of Likelihood function of p_A in (3.5) becomes

$$\ln(p_A) = \frac{5}{2} \ln(p_A) + \frac{1}{2} \ln(2) + \frac{5}{2} \ln(1 - p_A) \quad (4.1)$$

The graph of the function in (4.1) is shown in the following Figure.

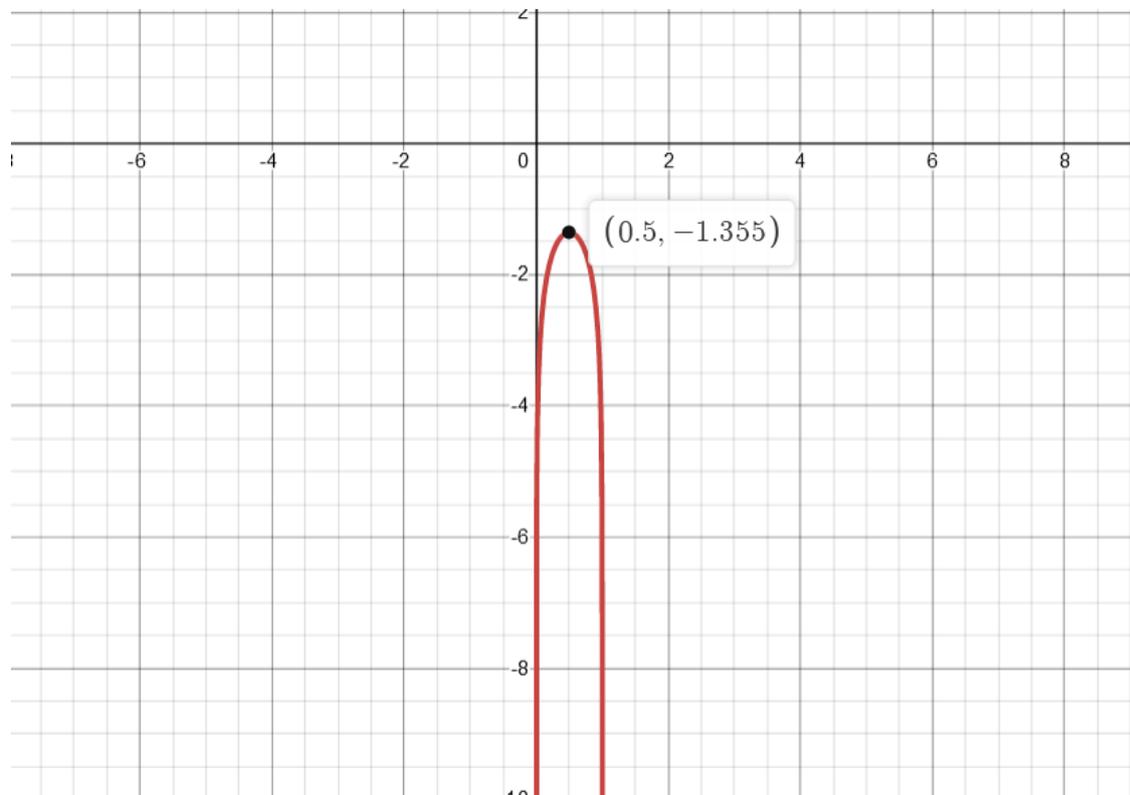


Figure 1: Graph of $\frac{5}{2} \ln(p_A) + \frac{1}{2} \ln(2) + \frac{5}{2} \ln(1 - p_A)$

From Figure 1, we notice that the function attains maximum at $p_A = \frac{1}{2} = 0.5$. Also for the values

$n_{AA} = 2n_{Aa} = n_{aa} = 1$ from (3.7) we find that $p_A = \frac{1}{2} = 0.5$ verifying the observation made through graph in Figure 1.

5. Conclusion

Considering two types of alleles of the form A and a , their population frequencies n_{AA}, n_{Aa}, n_{aa} and making use of Hardy – Weinberg Equilibrium principle, I arrived at three equations (3.1), (3.2), (3.3) with one unknown p_A .

Constructing the logarithm of likelihood function with respect to the three equations for the population frequencies, I had determined the maximum likelihood estimate. Such an estimate is given by the expression in (3.7). Using this value and $p_G = 1 - p_A$, I arrived at equation (3.8). Equations (3.7) and (3.8) provide the allele frequencies that we require, at the same time such frequencies maximize the likelihood functions constructed using their population frequencies. Finally, the maximum likelihood estimate values are verified with a suitable example in section 4. The ideas discussed in this paper provide us a thought of how alleles are distributed in the whole genotype population of a diploid organism.

REFERENCES

- [1]. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in B. N. Petrox and F. Caski, Second international symposium on information theory, pp.267-281.
 - [2]. Chen, J.J., The Hardy-Weinberg principle and its applications in modern population genetics, *Frontiers in Biology*, volume 5, pp. 348 – 353, (2010).
 - [3]. R. Sivaraman, Mathematical Modeling of Recovery Curves, *African Journal of Mathematics and Statistics Studies*, Volume 3, Issue 5, 2020, pp. 38 – 41.
 - [4]. R. Sivaraman, Markov Process and Decision Analysis, *Journal of Mechanics of Continua and Mathematical Sciences*, Volume 15, No. 7, July 2020, pp. 9 – 16.
 - [5]. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. *Molecular biology of the cell*. Garland, New York, 1989.
-