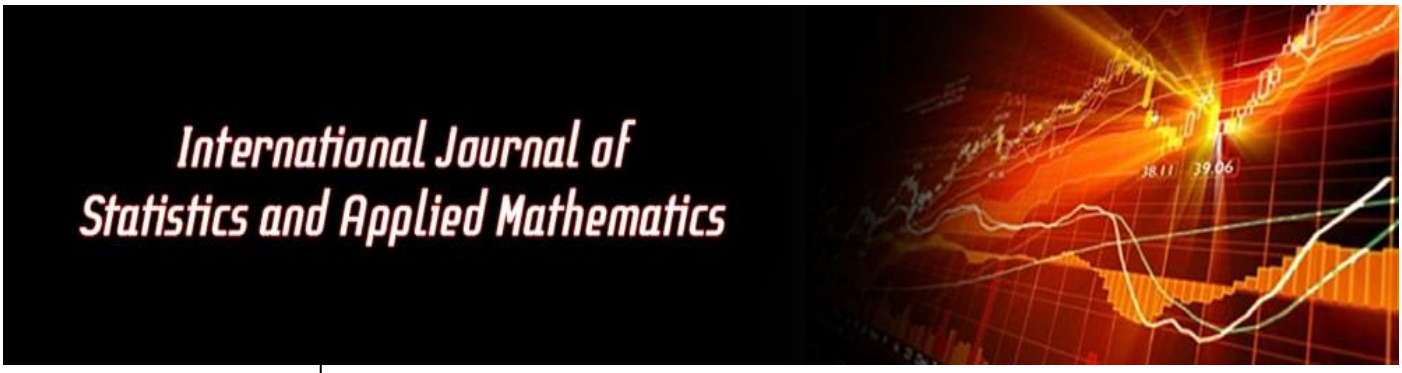
International Journal of Statistics and Applied Mathematics 2018; 3(1): 457-459



ISSN: 2456-1452
Maths 2018; 3(1): 457-459

www.mathsjournal.com
Received: 16-11-2017
Accepted: 24-12-2017

M Ramesh
Data Scientist, Development bank of Singapore (DBS), Singapore

P Vishnu Priya
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

G Madhusudan
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

P Srivyshnavi
Senior Assistant Professor, Department of CSE, SPMVV Engineering College, Tirupati, Andhra Pradesh, India

P Muralinadha Reddy
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

P Balasiddamuni
Rtd. Professor, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

Correspondence
G Madhusudan
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

Iterative procedure for estimating gene frequencies in genetics data analysis

M Ramesh, P Vishnu Priya, G Madhusudan, P Srivyshnavi, P Muralinadha Reddy and P Balasiddamuni

Abstract
Genetics is the mother of breeding and statistics is the only tool to analysis genetic data to arrive at conclusions which hold good in general with a high degree of accuracy and precision. Biometrical genetics is a branch of genetics that uses various statistical concepts and methods to study the genetical research problems. This may be divided into two major parts namely, population Genetics and Quantitative Genetics. Population genetics deals with the gene frequencies, genotypes and phenotypes in the Mendelian population.
The heredity units which are transmitted from one generation to the next generation are called 'Genes'. In the population genetic, individual genes have no meaning. It is the gene pool consists of population that matters. The gene pool consists of all the genes in all their allele forms in the reproductive gamates of the population. It is described as total genetic information possessed by a population which is available to next generation.
One of the basic Idea of population genetics is the Gene Frequency. If the gene frequencies of certain genes are estimated, then the expected genotypic frequency of the next generation can be calculated in a Random Mating.

Keywords: Iterative procedure, estimating gene frequencies

1. Introduction
Bio-statistical Techniques have a wide number of practical applications in Genetics and plant breeding.
One may define Biostatistics, Bio statistical Genetics, Quantitative Genetics and Population Genetics as follows:

a) Biostatistics or Biometrics: It is a branch of either Biology or Applied statistics which deals with the application of statistical techniques to study various research problems in the Biology.

b) Biostatistical genetics or biometrical genetics: It is a branch of Genetics that uses various statistical concepts and methods to study the genetical research problems sometimes it is known as statistical genetics or mathematical genetics. This may be divided into two major parts namely (i) Quantitative Genetics and ii) Population Genetics

c) Quantitative Genetics: It can be considered as a branch of Biostatistical Genetics which deals with the study of polygenic or quantitative characters.

The main features of Quantitative genetics are

i) Quantitative genetics is an extension of Mend Elian genetics firmly based on Mend Elian principles of heredity;
ii) It provides ways and means for the study of polygenic variation, which is not possible by Mend Elian genetics;
iii) It deals with continuous variation.
iv) Its estimates are based on means, variances and covariances;
v) It helps in better understanding of genetic principles and better planning of crop improvement programmes; and
vi) R.A. Fisher laid the foundation of quantitative genetics in years 1918.

i) Population Genetics: It is a branch of Bio statistical Genetics which deals with the Gene frequencies, and Genotypes and phenotypes in the mend Elian population. These populations have two main features namely a) Random mating and b) Equal survival of all genotypes.

The main features of population Genetics are

(i) It deals with oligogenic variation or discontinuous variation; (ii) Its estimates are based on frequencies and ratios; (iii) It cannot be used for analysis of polygenic variation; and (iv) Hardy and Weinberg laid the foundation of population genetics in year 1908.

2. Iterative Procedure for Estimating the Gene Frequencies

The estimation of population frequencies of alleles at a given genetic locus or the estimation of the population frequencies of Haplotypes is referred to as 'Gene Frequency Estimation'. A Haplotype, derived from the term 'Haploid Genotype' refers to the particular set of alleles present at a series of linked loci on a chromosome i.e., alleles at loci that are present relatively close together on a continuous strand of DNA. Each human being has two haplotypes for any given series of linked autosomal loci; one inherited maternally, and the other, paternally. The haplotype transmitted to an offspring of that individual may be identical to one of these two, or may reflect reshufflings due to recombination events. Allele and haplotypic frequencies may be estimated from various types of human data, including information from pedigrees, parent child dyads, or individuals; however, estimation based upon random samples of individuals is emphasized in this article.

Good estimates of gene frequencies are needed for a number of purposes. They are helpful in planning genetic studies. Marker allele frequencies are required for the application of certain types of linkage analysis.

If a random sample of N individuals is collected and there are X_i copies of allele Ai among the 2N alleles, then the estimator of the allele frequency q_i is $X_i/2N$, the maximum likelihood estimator. The variance of q_i can be obtained by using Binomial distribution as $\frac{q_i (1 - q_i)}{2N}$. Irrespect of the nature of the relationships the maximum likelihood estimation method may be used to estimate the gene frequencies.

Support that a random sample of individuals is taken from the relevant population and that there are k distinguishable phenotypes associated with the locus of interest indexed by k = 1, 2, ….., K. Usually, an assumption may be made in order to set up the correspondences between genotype and phenotype, expressed via the allele frequencies. By assuming the Hardy-Weinberg equilibrium law, the population frequency ϕ_k of each phenotype can be expressed as a function of the allele frequencies.

For n_k individuals with the k^{th} phenotype, the Likelihood function (L) of the sample is given by

$$L \alpha \prod_{k=1}^{K} \phi_k^{n_k}$$

The logarithmic likelihood function is given by

$$l_n, L = \sum_{k=1}^{k} n_k \, l_n \, (\phi_k) + constant$$

For instance, consider a locus with a coxominant alleles A_1, A_2, ….., A_i, ….., A_a plus a null allele A_x, which cannot be detected, with population allele frequencies q_1, q_2, ……., q_a, q_x.

One may have $\binom{k}{2}$ phenotypes of the type $A_i A_j$, which occur with population frequency zq_iq_j, and k phenotypes of the type that express only a single allele A_i, which may be either A_i homozygotes or heterozygotes involving a null allele and occur with population frequency ($q_i^2 + 2q_iq_x$), and a phenotype expressing no allele Ax, which occurs with population frequency q_x^2.

If one defines the sample counts of these phenotypes analogously, one may obtain the log likelihood.

$$l_n \, L = \sum\sum_{i<j} n_{ij} \, L_n \, (2q_iq_j) + \sum_i n_i L_n \, (q_i^2 + 2q_iq_x) + n_x \ln(q_x^2)$$

The estimates of the allele frequencies are obtained using standard iterative methods based upon the maximum likelihood method, and estimates of the standard errors can be derived numerically by using the Fisher's amount of information. In the application of iterative process any of the Bernstein estimates can be taken as the initial estimates. If f_i denotes the proportion of individuals expressing the i^{th} allele, one of the Bernstein estimator for q_i is

$$q_i^* = 1 - (1 - f_i)^{1/2}$$

q_x is generally estimated by subtraction.

3. Conclusions

Under the inferential aspects of Genetic parameters, one may consider the problems of estimation of genetic parameters, for instance, segregation ratios, Gene frequencies, Recombination values, Heritability coefficients, Repeatability coefficients, Genetic correlation coefficients and so on. Then one may turn to the examination of goodness of fit with the estimated genetic parameters and to the comparison of sets of data.

An Iterative procedure for estimating the Gene Frequencies has been proposed in the present research work. The maximum likelihood estimation methods has been used by taking Bernstein estimates as initial estimates for the iterative procedure.

4. References

1. Agarwal BL, Agarwal SP. Statistical Analysis of Quantitative Genetics New Age International Publishers, New Delhi, 2007.
2. Finney DJ. Scores for the estimation of genetic parameters, Biometrics. 1950; 6:221-227.
3. Haldane JBS. An exact test for randomness of mating. J Genet. 1954; 52:631-635.
4. Hayman BI. The analysis of variance of dialleles tables Biometrics. 1954b; 10:235-244.
5. Kempthorne O. An Introduction to Genetic Statistics, the Iowa State University Press, Ames, Iowa, U.S.A, 1957.
6. Mather K, Jinks JL. Biometrical Genetics", Chapman and Hall London, 1971.
7. Rao MG, Jain JP. Effect of non-normality on response to selection in large population. J Ind. sco. Agri. stat. 1980; 32:82-94.
8. Reeve ECR. The Variance of the Genetic Correlation Coefficient, Biometrics. 1955; 11:357-374.
9. Singh P, Narayanan SS. Biometrical Techniques in plant Breeding, 2nd Edition Kalyani Publishers, New Delhi, 2000.
10. Wright S. Systems of mating I-V Genetics. 1934b; 6:111-178.
11. Wright S. Correlation and Causation, Journal of Agriculture Research. 1921; 20:557-585.
12. Wright S. System of Mating-I, The Biometric Relation between parent and offspring, Genetics. 1921; 6:111-123.