Statistics and Mathematical Modelling;  
A Major Recent Modern Tool in Biotechnology and Bioinformatics Data Analysis

Suneetha V.*, Bishwambhar Mishra, Parul Kamat, Gopi Chand T., Saranya C., Rani Anupama and Alok Prakash

School of Bio Sciences and Technology
VIT University, Vellore-632 014, India

*Corresponding author
vsuneetha@vit.ac.in, bishwambharmishra@vit.ac.in

Copyright © 2013 Suneetha V. et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Statistics and Mathematical models can be in many forms, including differential equations, or theoretic models. Lack of correlation between theoretical mathematical models and experimental measurements sometime leads to development of better theories in the area of biology. The mathematical design finds wide application in nutrient media optimization for microbial enzyme production. On the other hand these mathematical tools in the field of statistics play an important role with data analysis of proteomics and genomics studies. For the MS/MS spectroscopic data, micro array, RT-PCR data analysis, linkage study, gene-gene interaction studies are very difficult to interpret, but the statistical tools made it very easier from the conventional method of data analysis in both Biotechnology and Bioinformatics.

Keywords: Mathematical modelling, Statistics Experimental measurement, Optimization Factors

INTRODUCTION

Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data and it deals with all aspects of this, includ-
ing the planning of data collection in terms of the design of surveys and experiments with different types of variables where as mathematical model usually deals with a system by a set of variables and a set of equations that establish relationships between the variables\[1, 2\].The variables represent some properties of the system like signals, counters, timing data and occurrence of events \[7, 9\].Media composition for the growth and production of the specific fermented product is an essential step which requires the selection of optimum nutritional and physical factors supporting the growth and production of desired product from the specific organism \[5\].A series of the experiments is performed in which the optimized results of the variable factor of the previous experiment is then incorporated in the next predefined experiment and the same procedure is adapted for all the other parameters to obtain complete optimization \[6\].

In the present review paper the various applications of the mathematical modelling and statistics for the process optimisation of both different types of parameters has been described.

**Data collection and Microbiological sampling plans**

Collection data (Data mining) is an important parameter of any type of research study. Inaccurate data collection sometimes impacts the results of a study which ultimately leads to invalid and false results \[6\].Generally in sampling of microbiological data the decision making process of a two-class plan is essentially defined by two numbers \[9\]. The first one determines the number of sample units that are to be drawn independently or randomly from the huge no of samples. The second one is the maximum allowable number of sample units yielding results of unsatisfactory test \[10,12\]. In situations where decisions are not based on results of presence-absence tests but on quantitative analytical results, three-class plans can be applied as an alternative to two-class plans working with data grouped according to a single microbiological limit \[5,6\].

**Mathematical modelling of complex biological systems and macroscopic tissue models**

It has been shown that how the mathematical tools with its various applications has been derived are applied for both in the modelling of complex biological systems constituted by different cell populations with \textit{in vivo} condition, which are carrier of a pathological state and to derive macroscopic tissue level equations \[3, 5\]. The complex biological systems all together with application of therapeutic treatment, which has been dealt with by various authors for less general models using both the diffusion ( in parabolic form ) or high field ( in hyperbolic form) \[7,8\].
Statistical analysis with special reference to optimize the bioinformatics data

Linkage studies

Linkage is the inherent tendency of different types of genes that are located proximal to each other on a chromosome to be inherited during meiosis process. The LOD score [Logarithm (base 10) of Odds], is a statistical tools to test for linkage analysis in all living systems [1, 9]. Moreover, the studies on human genetic variation deciphering the genetics of complex diseases (e.g., hypertension, hypotension, diabetes etc.) through genome-wide association and the various copy number variation studies[12].

Analysis of gene-gene interaction

Genes are always operating within the specific location of micro-environment of the cell, and expression of one gene can be influenced by presence or absence of other genes. In addition, an environmental conditions sometimes trigger the expression of a gene that in turn modifies other genes [11]. A single gene association model may be gene-gene and gene-environment interactions may pose greater challenges and also lead to more significant findings [8,10].

Estimation of DNA copy numbers

DNA copy number variation is a widespread and generally accepted common phenomenon among humans was firstly reported after the completion of the human genome project (HGP). Now a days the nutritional studies have also been conducted to provide vital clues for the cause of different forms of cancers, including colon cancer, prostate cancer and breast cancer etc [4, 6].

Mass spectrometry (Proteomics) data analysis with statistical tools

Mass spectrometry (MS) is an analytical tool that measures the mass-to-charge ratio of charged biological samples like peptides. This mass spectrum contain different types of peaks which is corresponding to the m/z values of different types of peptides in the original crude sample [9]. Generally, the region of m/z values chosen to be passed through is one of the largest peaks in the MS scan data, but most of the instrument will also allow the experimenter to program specific rules for the performance [10]. Thus, the main data structure of a tandem MS/MS run is alternating scans, beginning with ion concentrations from the original sample in one scan (MS) and then the ion concentrations from the tandem MS/MS scan [6, 11,12].
Statistical methods to analyse the microarray experiments

A DNA microarray is a collection of microscopic DNA spots attached to an inert solid surface. Each DNA spot contains picomoles of a specific DNA sequence, known as probes. These can be a short fragment of a gene or other genetic element like DNA that are used to hybridize with a complementary DNA or RNA (cDNA or cRNA) known as target in highly stringent conditions [2, 5]. For the small sample sizes analysis the microarray data sets can make the normality assumption a good choice [8, 12].

CONCLUSION

In this review, we have focused on such application of statistics and mathematical models and their specific application in genomics, proteomics and bioinformatics studies. Statistical analysis is a necessary tool to test hypotheses in some biomedical research also. The calculation of sample size is a critical steps in designing the high throughput studies like microarray analysis and MS/MS studies.

ACKNOWLEDGEMENT

The authors want to acknowledge the DST, India and VIT University, Vellore for good infrastructure and smart laboratory facilities to carry out this work.

REFERENCES


Received: January, 2013