Int. J. Open Problems Compt. Math., Vol. 3, No. 3, September 2010
ISSN 1998-6262; Copyright © ICSRS Publication, 2010
www.i-csrs.org

Bioinformatics Computing of Parametric Information Connected with Industrially Important Bacteria

Sant Saran Mishra

Department of Mathematics and Statistics, Dr Ram Manohar Lohia Avadh University, Faizabad-224001, U P, India.
e mail: sant_x2003@yahoo.co.in

Abstract

In this paper, a parameter-information and its minimum variance have been focused as performance indicators for the growth and decay of the industrially important bacterial population. A statistical inference technique with its computing algorithm has been employed to numerically demonstrate the use of the model under consideration as an innovative aspect of the bioinformatics computing related to bacterial population.

1 Introduction

Information theory is a branch of applied mathematics and engineering. This science involves the measurement and quantification of information which is used in various fields such as statistical inference, networks communication, molecular codes model selection in ecology and neurobiology with quantum computing. Thus, we can say that the field is at the intersection of mathematics, statistics, computer science and biology etc which perhaps a computing science requires this kind of intersection. Information theory is generally considered to have been founded in 1948 by Claude Shanon in his work a mathematical theory of communication having key paradigm of information over noisy channels. It is closely related with cybernetics, computing, complexity science, informatics, machine learning and system sciences etc. Likelihood function is a function of the parameter of a statistical model that has a prominent role in the theory of statistical inference. In non-technical parlance, likelihood is usually a synonym for probability but in statistical usage there is a clear distinction. Probability tells us to predict unknown ***outcomes*** based on known ***parameters***, whereas likelihood tells us to estimate unknown ***parameters*** based on known ***outcomes***. Reversing this concept, we can construct the likelihood function: given outcome **A**, use the likelihood function **L** (**B**|**A**) to reason about parameter **B**. Formally, a likelihood function is a conditional probability function. A W F Edwards [5] defined support as the natural algorithm of likelihood ratio and support function is defined as the natural logarithm of the likelihood function and score function is defined as the partial derivative of likelihood function with respect to population parameter. The Fisher information is a way of measuring the amount of information that an observable random variable carries about an unknown population parameter upon which likelihood function depends. The likelihood function is the joint probability of the data, the ***random variables***, conditional on the value of parameter, ***as a function of it***. Since the expectation of the score is zero, the variance is simply the second moment of the score, the derivative of the log of the likelihood function with respect to parameter. Hence the Fisher information can be written as the expectation with respect to the observations X, of the squared score (vide Savage [10]). A random variable carrying high Fisher information implies that the absolute value of the score is often high. The Fisher information is not a function of a particular observation, as the random variable **X** has been averaged out. The concept of information is useful when comparing two methods of observing a given random process. Thus Fisher information is also defined as the negative of the expectation of the second derivative of the log of ***pdf*** with respect to θ. Information may thus be seen to be a measure of the sharpness of the support curve near the maximum likelihood estimate of θ. In a simple language, we can say that low expected second derivative will give us low information and while a sharp one having a high expected second derivative will give us high information. The Cramer-Rao inequality states that the inverse of the Fisher information is a lower bound on the variance of any unbiased estimator of population parameter θ. Bacteria are very useful for our life and scientific development (vide Lesk [1] and Marketa and Baum [13] as follows:

For food, microorganisms are used in brewing, winemaking, baking, pickling and other food-making processes.They are also used to control the fermentation process in the production of cultured dairy products such as yogurt and cheese.

For energy, microbes are used in fermentation to produce ethanol, and in biogas reactors to produce methane. Scientists are researching the use of algae to produce liquid fuels, and bacteria to convert various forms of agricultural and urban waste into usable fuels.

In science, microbes are also essential tools in biotechnology, biochemistry, genetics, and molecular biology. The yeasts (Saccharomyces cerevisiae) and fission yeast (Schizosaccharomyces pombe) are important model organisms in science, since they are simple eukaryotes that can be grown rapidly in large numbers and are easily manipulated. In the field of warefare, in the middle ages, diseased corpses were thrown into castles during sieges using catapults or other siege engines. Individuals near the corpses were exposed to the deadly pathogen and were likely to spread that pathogen to others.

In the health sector, microorganisms can form an endosymbiotic relationship with other, larger organisms. For example, the bacteria that live within the human digestive system contribute to gut immunity, synthesise vitamins such as folic acid and biotin, and ferment complex indigestible carbohydrates.In the sector of immunology and diseases control, microorganisms are the cause of many infectious diseases. The organisms involved include pathogenic bacteria, causing diseases such as plague, tuberculosis and anthrax; protozoa, causing diseases such as malaria, sleeping sickness and toxoplasmosis; and also fungi causing diseases such as ringworm, candidiasis or histoplasmosis. decomposition required to cycle nitrogen and other elements back to the natural world. Also hygiene is the avoidance of infection or food spoiling by eliminating microorganisms from the surroundings. As microorganisms, particularly bacteria, are found practically everywhere, this means in most cases the reduction of harmful microorganisms to acceptable levels. For example, a food sample might be analyzed on three different nutrient mediums designed to indicate the presence of "total" bacteria (conditions where many, but not all, bacteria grow), molds (conditions where the growth of bacteria is prevented by e.g. antibiotics) and coliform bacteria (these indicate a sewage contamination). Long and Pace [3] and Kussell et.al [4] have discussed various diversity analysis of bacteria in differential survival environments. Shuler et.al [11] presented a mathematical model for the growth of single bacteria. They also presented the results of investigation on the effect of the mean and variance of the individual lag time and the initial cell number on the mean and variance of the population lag time. A mathematical model for the growth process of the bacterium Bacillus subtilis is described by Attai [12]. The model is a highly structured one. The driving motivation for development of the model and explicit accounting of major interactions of metabolic networks in the model is related to our eventual goal that the model will be used in the analysis of complex biological patterns. Kenneth and Kamau [9] reviewed the history of growth theories by showing how classical growth models may be derived as special cases of a generic growth rate equation. They used linear combinations of growth and survival models to represent complex growth/survival curves and give practical examples utilizing nonlinear regression analysis. Often, bacteria are very small in the order of micrometers (10-6 meter) in the length. 1 μ m length contains about 1000 bacteria, placed end to end in one milimeter length like a width of a pencil line. The growth and decay of bacteria is very simple in terms of mathematics that is governed by the binary fission. It shows that original cell doubles it self and diviedes into two genetically identical cells. Before fission, DNA replication takes place and each cell gets complete genome from prestine one and therefore called daugther cells. This phenomenon of growth and decay of inoculum of bacteria positions an inquisitive quest related to its parametric information that can be exploited for their optimal utilization for industrial and scientific developments.

Quantification of low and high parametric information is believed to result into high and low sensitivity of estimator's variance which is meant for estimating the population parameter connected with industrially important bacterial population (vide Stephen [17] and Schervish [18]. Extensively reviewed literature corroborate that we sufficiently lack the significant work in the bioinformatics computing of parametric information connected with industrially important bacteria (vide Jean and Cedric [8]). With this objective in the mind, we propose in this paper a bioinformatics computing of parametric information connected with growth and decay of industrially important bacterial population as a novel endeavour in this field. An elegant mathematarical modelling and then computing algorithm related to parametric information of bacterial population have been developed and search program has been executed to numerically compute the performance indicators for exhibiting the use of the model under consideration. This paper has been organized in various sections such as introduction, model development and its analysis, computing algorithm and its tabular and graphic presentation, with conclusion.

2 Model Development and its Analysis

3 Development of Computing Algorithm and Numerical Demonstration

4 Open Problem

In the development and analysis of bacterial population model; growth, decay and life-cycle indicators are very important for predictive findings in this sort of investigation. We have so far attempted deterministic and inference approaches for such kind of model and bioinformatics computing was conducted for efficiently analyzing and demonstrating the model under consideration. Still, the problems connected with bacterial growth in fuzzy as well as in neural networking environments are open for further researches making the model more efficacious than previous ones because soft computing based on fuzzy and neural network systems is more versatile and productive than deterministic and probabilistic environments.

5 Conclusion

We have computed an informatics parameter as an efficient apparatus of sensitivity because non-parametric information has advantage of being distribution independent as well as insensitive to extreme values. The disadvantage of non-parametric information is the complexity, lower power and time required for computation. In contrast, parametric information is simple and easy to compute but rely upon the assumption of a “Gaussian” distribution, vide graphs III and IV. Parametric information is known to be generally robust even when the assumption of Gaussian distribution is violated. Moreover, I have been able to analyze and compute MVU estimator as a versatile indicator because it is contingent upon theory of sufficient statistics. The first step in deriving the MVU estimator is to obtain the sufficient statistic for the given probability density function which is assumed to be a function of observation vector such that the conditional probability density function of the observation vector does not depend on the unknown estimation parameters forms a sufficient statistic for estimation of the parameter.

References

[1] Arthur Lesk, *Introduction to Bioinformatics*, Oxford University Press, USA, (2008), 06-02.

[2] Anders Hald, "On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares", **Statistical Science 14** (2), (1999).

[3] De Long E and Pace N, "Environmental diversity of bacteria and archaea", *Syst Biol* **50** (4): 470–8, (2001).

[4] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler, "Bacterial Persistence: A Model of Survival in Changing Environments", **Genetics 169**, (2005), 1807-1814

[5] A.W.F. Edwards, **Likelihood**, Johns Hopkins University Press, Baltimore, (1992).

[6] A. ***A***. Hald, **History of Mathematical Statistics from 1750 to 1930**, New York, Wiley, (1998).

[7] John W. Pratt. "F. Y. Edgeworth and R. A. Fisher on the Efficiency of Maximum Likelihood Estimation". **The Annals of Statistics 4** (3), (1976), pp. 501-514.

[8] Jean-Michel Claverie and Cedric Notredame, Bioinformatics, For Dummies, (2006), 12-18.

[9] Kenneth M. Pruitt and David N. Kamau, "Mathematical models of bacterial growth, inhibition and death under combined stress conditions", *Journal of Industrial and Microbiology and Biotechnology*, Vol. 12, Springer Berlin/Heidelberg, (1993).

[10] Leonard J. Savage. "On Rereading R. A. Fisher". **The Annals of Statistics 4** (3), (1976), pp. 441-500.

[11] M. L. Shuler, S. Leung, C. C. Dick, "A Mathematical model for the growth of a single bacterial cell", *Wiley Interscience*, Vol 326, Biochemical Engineering, (2006), pp35-52.

[12] M. M. Ataai, "A Mathematical Model for examining growth and sporulation processes of Bacillus subtilis Chemical and Petroleum Engineering Department", University of Pittsburgh, Pittsburgh, Pennsylvania 15261, (1989).

[13] Marketa Zvelebil and Jeremy Baum, *Understanding Biomathematics*, Garland Science, (2007), 08-29.

[14] S.S. Mishra, V. Ram Tripathi, D.C. Shukla and S.K. Sharma, "Bioinformatics computing of growth of industrially important bacteria", *International Journal of Information Science and Computer Engineering*, Communicated, (2009).

[15] C. Neil, Jones Pavel and A. Pevzner, *An Introduction to Bioinformatics Algorithms*, The MIT Press, (2004), 08-01.

[16] S.S. Mishra and D.K. Yadav, "Cost and proft analysis of Markovian queueing system with two priority classes: A computational Approach", *International Journal of Applied Mathematics and Computer Sciences*, 5:3, (2009), pp 150-156.

[17] Stephen M. Stigler. "Francis Ysidro Edgeworth, Statistician". *Journal of the Royal Statistical Society. Series A (General) 141* (3), (1978).

[18] Schervish Mark J. *Theory of Statistics*. New York: Springer. Section 2.3.1., (1995).