Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU24064

Inter-Discipline Between Mathematical Modeling, Statistical Inference and Biology

Yuliang Han

Wuhan Britain-China School, Wuhan, China
275819787@qq.com

Abstract: This paper mainly studies the interdisciplinary applications of mathematics and statistics considering biological systems and their defining attributes with special attention to their role in ecological and epidemiological modeling. It outlines how the exponential and logistic models, along with the Lotka-Volterra systems, reproduce the population and species interaction dynamics on a mathematical level. The importance of compartmental models in infectious disease modeling is also discussed, explaining what makes SIR so crucial to the process. Then, it provides examples including statistical estimation of parameters, such as least squares, maximum likelihood, and Bayesian reasoning to explain the observation theoretically verged with actual reality. This work as a whole illustrates the explanatory and practical rigor and formality applied alongside biological issues tackled through mathematics and statistically advanced techniques.

Keywords: mathematical biology, epidemiological modeling, statistical biology, regression analysis

1. Introduction

Using mathematics in biology has transformed the life sciences in the last century by shifting from a qualitative, observable approach to computable prediction techniques. Descriptive narratives formed the foundation of previous ecological and epidemiological research, which now requires decomposition, simulation, and forecasting of mathematical biological systems—referred to as systems biology—within the confines of mathematics. The inception of biostatistics, with its focus on model building through parameter estimation—least squares, maximum likelihood, and Bayesian inference—enables calibrated models to correspond with observational data. These tools surpass mere description of intricate systems; they design, guide policy, and build frameworks for understanding systems. Moreover, applying a mathematical structure to empirical data establishes a feedback loop capable of creating models that determine experimental design and outline essential data collection processes. Such design ensures biological relevance while creating insights into the phenomena being studied and the pace of discovery is accelerated. Biology, with the help of models, can provide new unparalleled accuracy, ensuring that the models have unparalleled reliability.

The rest of the paper is organized as follows: Chapter 2 introduces the mathematical models applied in biology; several classic models are discussed here. Chapter 3 emphasizes the importance of statistical algorithms applied in biology and gives examples of how to use these algorithms to solve the real-world parameter estimation problem.

DOI: 10.54254/2753-8818/2025.AU24064

2. Inter-discipline between mathematics and biology

2.1. Background

The fusion of mathematics and biology marks the first occurrence of the intersection between disciplines. The novelty of this cross disciplinary blend resulted in a fundamental shift in the understanding of ecological systems, transforming the once-reliant observational biology into a predictive science. Over this past century, rest of the ecological and epidemiological research also underwent the same transformation.

The foundation of early population studies relies on description-based natural history framework and is devoid of any propelling potential to address urgent problems like conservation, pest outbreak, and infectious disease control.

To manage that, single-species population models have become increasingly important, aiding the transitioning shift toward quantitative modeling, wherein existing biological logic is turned into conjectures that can be tested.

These models attempt to explain the synergetic relationship between the internal growth rates, constraints posed by the environment, and random disturbances. For example, Ludwig et al.'s study of spruce budworm outbreaks demonstrated how nonlinear feedback of forest foliage and herbivore density could explain the recursive cycle of infestations—not something reducible to verbal theories alone [1]. Turchin's synthesis of theoretical and empirical population dynamics also presented the important principle of complex and rigorous mathematics intertwined with ecology [2]. Meanwhile, Murray points out that the equations in question must find that middle ground between realism in biology and ease in mathematics, which pervades the entire background of Mathematical Biology [3].

2.2. Continuous growth models for single species

The focus on single-species models signified the evolution of descriptive to mechanistic ecology. Earlier ecologists understood that practical problems such as fisheries management and pest outbreaks required more than simple qualitative observation. Murray highlights how even such simple equations as the logistic growth model surfaced not as hypothetical exercises, but as real biological problems awaiting solutions [3].

Population models incorporating more than one generation, overlapping generations, and constant reproduction tend to perform better using differential equations. One can start with the exponential growth equation, which seems to conflict with the concept of finite resources provided by nature.

$$\frac{dN}{dt} = rN \tag{1}$$

However, this model is counterintuitive since resources are assumed to be infinite, which is clearly biologically impossible.

But as Ludwig et al. demonstrated in their studies on resource management, although this formula cannot be applied in reality, it could serve as a significant null hypothesis. Specifically, it is regarded as a fundamental rule for further studies about realistic density-dependent effects [1]. Consequently, refinement to logistic growth model occurred subsequently. The parameter carrying capacity K is introduced to the equation. The example above demonstrates how mathematical models can evolve iteratively with the introduction and observations of biological reality.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{2}$$

DOI: 10.54254/2753-8818/2025.AU24064

2.3. Discrete-time models for single species

This begins to unfold as we analyze the different modeling approaches: discrete vs. continuous. The discrete-time difference equations are critical to model species that have distinct non-overlapping generations or periodic seasonal cycles of reproduction. For species such as rabbits who have pulsed breeding seasons, difference equations such as

$$N_{t+1} = rN_t\left(1 - \frac{N_t}{K}\right) \tag{3}$$

Align best with biological rhythms. In this equation, N_t denotes the population size in the generation n, r is the intrinsic growth rate, and K is the environment's carrying capacity. This model captures density dependent population regulation which slows down growth as approaches inflection point [3].

These discrete models exemplify naturally the annual field surveys where populations are counted at particular intervals. Regardless, discrete models seem to show unique behaviors that are absent in their counterpart continuous models. A particular insight from Turchin underscores the rabbit model's ability to generate complex dynamics, from stable balance to fluctuations when increases exceed the point of equilibrium [2]. This emphasizes the notion that uncomplicated rules in biology can lead to sophisticated patterns mathematically.

As Turchin mentions, this strength becomes a limitation when continuously reproducing organisms come into play. The "tyranny of time steps" of difference equations begins to cloud intergenerational processes such as slow consumption of resources, demanding alternative solutions for bacteria or human populations [2]. In the meantime, discrete models fail to represent continuous processes such as gradual resource depletion or age-stratified mortality.

2.4. Species interaction

The progression from single-species to multi-species models demonstrates how mathematics can assist biological field. An instance is Predator-prey systems, which is formalized by Lotka-Volterra equations:

$$\begin{cases} \frac{dP}{dt} = \beta P - \gamma PQ \\ \frac{dQ}{dt} = \alpha PQ - \delta Q \end{cases} \tag{4}$$

The above equations illustrate the preys N grow exponentially in the absence of predators P, and on the other hand, predators decline without prey [3]. This clearly shows how abstract biological concepts can be transformed into precise rules through mathematical modeling.

By adding adaptive foraging, Křivan expands this model. Predators who change prey types based on type and availability stabilize systems that classical models predict to be in constant decline [4]. This adaptability, formalized through optimal foraging theory, satisfies Murray's anticipations, Metcalf's goals, and other flexible food web theorists [3]. This is an example that demonstrates refinement based on formalized biological conjectures can only be achieved through mathematical formalization.

2.5. SIR model

In the area of epidemiology, the SIR model divides a population into three distinct groups: Susceptible S (those at risk of infection), Infected I (able to spread the infection), and Recovered R (who are immune or have died). This model was developed by Kermack and McKendrick to provide

DOI: 10.54254/2753-8818/2025.AU24064

a framework for understanding the dynamics of infectious disease spread [5]. The framework of the model is described as follows:

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = \beta SI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (5)$$

where β and γ represents the transmission rate and the recovery rate respectively. This model can be used by determining epidemic thresholds $R_0 = \beta/\gamma$ [5]. In real life, this parameter applies to guide public health interventions. Additionally, studies by Anderson and May's illustrate how mathematical modeling combining biological data can work for life-saving such as analyzing HIV [6].

All these advances rest on a dramatic change in understanding: biology is now constructed through the lens of mathematics.

3. Inter-discipline between mathematics and statistics

3.1. Background

Statistics is one of the most important key technologies in depicting the patterns of biological species evolution or the spread and outbreak of infectious diseases. A common scenario involves knowing the changes in species population over the past period or the number of infected individuals in a region during a specific time. The goal is to use these early observational statistics to predict future trends in species population dynamics or disease outbreaks. Parameter estimation is one of the fundamental problems in statistical inference, which needs to apply the known data observed from experiments to mathematical models. There are two main methods of parameter estimation: point estimation and interval estimation. The concept of point estimation is to construct statistics (e.g., sample mean or variance) as a single estimate of the parameter, while interval estimation provides a range of plausible values for the parameter within a specified confidence level. In this paper, point estimation is focused on. The most common methods of point estimation are the following methods: the least squares methods, the maximum likelihood estimation, and the bayesian estimation.

3.1.1. Least Squares Methods

Least Squares Methods (LSM) calculates the direct distance between the existing data and the models. As a core tool for parameter estimation, LSM has had a profound impact on biostatistics. Its application background can be traced back to the quantitative analysis needs of biological data and has continued to deepen with the advancement of modern biological experimental technologies, like fitting allometric slopes through measurement data [7], identifying the dynamic interactions in biochemical networks [8] and the analysis of fluorescence data [9]. The details of LSM algorithm will be introduced in the later chapter.

3.1.2. Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is proposed by R.A. Fisher in 1912. The core idea of MLE is that the parameter should maximize the probability of the occurrence of the observed data. MLE is based on the frequentist probability perspective; parameters are considered fixed but unknown, and are estimated by optimizing the likelihood function. MLE is still popular among the studies today, like estimating the biological growth variables [10] and phylogeny estimation [11].

DOI: 10.54254/2753-8818/2025.AU24064

3.1.3. Bayesian estimation

Proposed by Thomas Bayes and later refined by Laplace, the core idea is Bayes' Theorem:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \tag{6}$$

Here, $P(\theta)$ is the prior distribution, $P(X|\theta)$ the likelihood function of distribution X based on parameter θ, and $P(\theta|X)$ is the posterior distribution. Advances in computational methods, like Markov Chain Monte Carlo (MCMC) algorithms, enabled Bayesian approaches to handle complex models. Bayesian Estimation can be used in the estimation for dynamical models in systems biology [12] and conservation biology [13].

3.2. Application of point estimation methods

In this chapter, the simple linear regression (SLR) model is taken as an example to illustrate the application of MLE and Bayes estimation.

The simple linear regression model has broad applications in biomathematical modeling. In biostatistics, correlation analysis and regression analysis are related but distinct. Correlation investigates whether there is a relationship between variables and the degree of that relationship. Once a correlation is confirmed through correlation analysis, a functional relationship is then sought to describe the connection between the variables, which is regression analysis. The function constructed to express this relationship is called a regression equation.

The simplest regression equation is the simple linear regression equation. It is used to analyze the linear relationship between an independent variable X and a dependent variable Y. The SLR model is widely used, and an example is given as follows. Suppose the population size $N(t)$ of a species follows the Malthusian growth law:

$$\frac{dN(t)}{dt} = rN(t) \tag{7}$$

With the initial condition $N(0) = N_0$, the solution is $N(t) = N_0 \exp(rt)$. This shows that the population size at any time is entirely determined by the initial population size N_0 and the intrinsic growth rate r. The estimation of these two parameters can be transformed into the following linear regression equation:

$$Y = \beta_0 + \beta_1 t \tag{8}$$

Where $Y = \ln(N)$, $\beta_0 = \ln(N_0)$ and $\beta_1 = r$. Formula (10) fits the SLR model perfectly. In reality, observasion error is always introduced into the model, it can be formulated as:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{9}$$

4. Conclusion

Considering mathematics and statistics as just an analytical tool is a very shallow way of perceiving them, since they serve to provide context that further enhances understanding of a biological system. To illustrate this, numerous attempts have been made to depict biological phenomena using mathematical models, such as the dynamics of population growth models starting from individual species through to more advanced models like predator-prey modeling systems or SIR compartmental modeling systems for infectious diseases. Such models become deterministic when paired with advanced empirical observation estimation techniques such as the least squares method, Bayesian inference, Maximum Likelihood Estimation, or other sophisticated approaches and undergo model verification and model validation, thereby bridging the gap between models and reality.

DOI: 10.54254/2753-8818/2025.AU24064

While a lack of data prevents the use of advanced statistical methods, oversimplified models in biology can hide critical components. The intertwining of biology, mathematics, and statistics has significant potential to enhance predictive capabilities and biological comprehension, which is crucial for public health and conservation efforts.

The combination of detailed real-time monitoring systems with innovative evaluation technologies creates unmatched opportunities for the management of entire ecosystems and the control of epidemic outbreaks. These collaborations usher in a thrilling period of swift progress in many disciplines, showcasing the future of biomathematics as the creation of advanced stochastic models with both discrete and continuous elements in relation to biological systems. At the end of the day, the interaction between evolving theories and plentiful data will give rise to new insightful biological discoveries.

References

[1] *Ludwig, D., Jones, D. D., & Holling, C. S. (1978). Qualitative analysis of insect outbreak systems: The spruce budworm and forest. Journal of Theoretical Biology, 72(3): 481–503.*

[2] *Turchin, P. (2003). Complex population dynamics: A theoretical/empirical synthesis. Princeton University Press.*

[3] *Murray, J. D. (2002). Mathematical biology: I. An introduction (3rd ed.). Springer.*

[4] *Křivan, V. (2008). Adaptive foraging and flexible food web topology. Ecology Letters, 11(12):1317–1326.*

[5] *Kermack, W. O., McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772): 700–721.*

[6] *Anderson, R. M., & May, R. M. (1991). Infectious diseases of humans: Dynamics and control. Oxford University Press.*

[7] *Kilmer, J. T., & Rodríguez, R. L. (2017). Ordinary least squares regression is indicated for studies of allometry. Journal of evolutionary biology, 30(1): 4-12.*

[8] *Kim, J., Bates, D. G., Postlethwaite, I., Heslop-Harrison, P., & Cho, K. H. (2007). Least-squares methods for identifying biochemical regulatory networks from noisy measurements. BMC bioinformatics, 8, 1-15.*

[9] *Straume, M., Frasier-Cadoret, S. G., Johnson, M. L. (2002). Least-squares analysis of fluorescence data. In Topics in Fluorescence Spectroscopy: Principles. Boston, MA: Springer US, pp. 177-240.*

[10] *Srivastava, A., Saini, S., Ding, Z., Grenander, U. (2005). Maximum-likelihood estimation of biological growth variables. In Energy Minimization Methods in Computer Vision and Pattern Recognition: 5th International Workshop, EMMCVPR 2005, St. Augustine, FL, USA, November 9-11, 2005. Proceedings 5. Springer Berlin Heidelberg, pp. 107-118.*

[11] *Huelsenbeck, J. P., Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and systematics, 28(1): 437-466.*

[12] *Linden, N. J., Kramer, B., Rangamani, P. (2022). Bayesian parameter estimation for dynamical models in systems biology. PLoS computational biology, 18(10), e1010651.*

[13] *Wade, P. R. (2000). Bayesian methods in conservation biology. Conservation biology, 14(5):1308-1316.*