Proposal for IB 201/IB 120: Introduction to Quantitative Methods In Biology

Course Description: This course provides a fast-paced introduction to a variety of quantitative methods used in biology and their mathematical underpinnings. While no topic will be covered in depth, the course will provide an overview of several different topics commonly encountered in modern biological research including differential equations and systems of differential equations, a review of basic concepts in linear algebra, an introduction to probability theory, Markov chains, maximum likelihood and Bayesian estimation, measures of statistical confidence, hypothesis testing and model choice, permutation and simulation, and several topics in statistics and machine learning including regression analyses, clustering, and principal component analyses. The course includes a lab section focusing on building student skills in modern computational methods for biological data analysis.

Expected student learning outcomes: One important pedagogical goal of the course is to serve as a refresher on mathematics and statistics for students from the biological sciences. In addition, students will acquire advanced techniques for quantitative analyses in the biological sciences.

Finally, students will learn hands-on skills on data analysis through the lab section.

The specific learning outcomes of the course are as follows:

- Familiarity with basic differential equations and their solutions.
- Ability to model simple relationships between biological variables using differential equations.
- Ability to manipulate matrices using multiplication and addition.
- An understanding of powers of matrices and the inverse of a matrix.
- Familiarity with the use of matrices to model transitions in a biological system with discrete categories.
- An understanding of basic probability theory including some of the standard univariate random variables, such as the binomial, geometric, exponential, and normal distribution, and how these variables can be used to model biological systems.
- Ability to calculate probabilities of discrete events using simple counting techniques, addition of probabilities of mutually exclusive events, multiplication of probabilities of independent events, the definition of conditional probability, the law of total probability, and Bayes' formula, and familiarity with the use of such calculations to understand biological relationships.
- Ability to calculate means and variances for a sample and relate it to expectations and variances of a random variable.
- An understanding of sampling and sampling variance.
- Ability to classify states in discrete time Markov chains, and to calculate transition probabilities and stationary distributions for simple discrete time, finite

state-space Markov chains, and an understanding of the modeling of evolutionary processes as Markov chains.
- Ability to define likelihood functions for simple examples based on standard random variables.
- An understanding of the principles used for point estimation, hypothesis testing, and the formation of confidence intervals and credible intervals.
- Ability to implement simple statistical models in R and to use simple permutation procedures to quantify uncertainty.
- Familiarity with covariance, correlation, ordinary least squares, and interpretations of slopes and intercepts of a regression line.
- Ability to implement standard and logistic regression models with multiple covariates in R.
- Familiarity with the assumptions of regression and methods for investigating the assumptions using R.
- Familiarity with random effects models and ability to implement them in R.
- Familiarity with analysis of variance and ability to implementation it in R.
- Familiarity with principal components analysis, other methods of clustering, and their implementation in R.
- Familiarity with one or more methods used in machine learning/statistics such as hidden Markov models, CART, neural networks, and/or graphical models.
- Ability to carry out various procedures for data visualization in R.
- Familiarity with functional programming in R and/or Python and ability to define new functions.
- Ability to work in a Unix environment and manipulating files in Unix.
- Familiarity with python allowing students to understand simple python scripts.

Tentative week-by-week schedule:

Week 1: Math reminders on calculus of one variable and special functions (logarithmic and exponential function, limits, differentiation, integration, series).
Lab: Introduction to Unix.

Week 2: Differential equations I (first order ODEs, examples of exponential growth, logistic growth)
Lab: Introduction to R. Importing data. Data frames. Basic operations.

Week 3: Differential equations II (coupled differential equations, Lotka-Volterra, more)
Lab: Data visualization in R.

Week 4: Linear algebra (matrix addition and multiplication, order of operations, powers of a matrix, Leslie Matrix as example, identity matrices, determinants, inverse of a matrix, eigenvalues and eigenvectors).
Lab: Using R for matrix manipulation and other math problems. Modeling of age structured populations as example.

Week 5: Introductory probability theory I (axioms of probability, conditional probability, Bayes' formula, discrete random variables, Binomial, Poisson).
Lab: Introduction to Python I. IDLE. Editing Python scripts. Importing data. Variables.

Week 6: Introductory probability theory II (expectation, variance, continuous random variables, exponential distribution, normal distribution and its properties).
Lab: Introduction to Python II. Strings, Lists, and Tuples. If-Else. Loops. Functional programming in Python

Week 7: Markov chains (discrete time Markov chains, transition probabilities, classification of states, stationary distribution, examples from evolutionary biology).
Lab: Introduction to BioPython. Accessing external data.

Week 8: Estimation, bias, method of moments estimation, maximum likelihood estimation, Bayesian estimation.
Lab: Simple statistics in R. More on data-frames and on accessing and extracting data in R.

Week 9: Confidence and credible intervals, hypothesis testing, model choice.
Lab: R exercise on constructing likelihood functions, optimizing the functions, and performing likelihood ratio tests.

Week 10: Regression analysis.
Lab: Linear regression in R using lm. Visualizations of data and residuals.

Week 11: Regression analysis continued. Permutation, simulation, and bootstrap
Lab: Linear regression in R continued. Random forests, Using permutation in R.

Week 12: Clustering and PCA.
Lab: PCA. K-means. Hierarchical Clustering. Visualization of clustering using dendrograms.

Week 13: Select topics in machine learning (HMMs, elimination algorithms, neural networks or other topics).
Lab: Select advanced topics using R.

Course requirements: Submission of weekly lab report to GSI. One midterm and a final exam.

Course readings: The course will be based on lecture notes developed by the instructors. Lecture notes for each week will be posted on bCourses.

Grading and weight of requirements: 50% lab reports. 20% midterm. 30% final exam. Grading will be done on a curve independently for undergraduate and graduate students.

GSI: The course will have one 50% GSI who is responsible for teaching and grading the lab section.

Prerequisites: Graduate standing or Biology 1A, Biology 1B, a course in statistics such as Data 8, Stat 2 or Stat 20, and two semesters of college level math including calculus such as Math 10A and Math 10B. Undergraduate students engaged in honors research, or other supervised research, are preferred. Graduate students who have not previously taken any statistics courses may benefit from taking Data 8, Stat 2 or Stat 20, or similar courses, before enrolling. . Previous knowledge of R is not necessary.

Justification for cross listing: While the primary goal of the class is to provide an introduction to quantitative methods for graduate students from the biological sciences, it will also, similarly, provide a suitable introduction for undergraduate students who are engaged in undergraduate research, intend to move into biological research, and/or are working to prepare themselves for graduate school in the biological sciences.

Units and hours: The class is approved for 4 units with 3 hours of lecture and 3 hours of lab per week.