Course OPTIMIZATION AND DATA ANALYSIS IN BIOLOGY

Basic Information

This is a course, which contributes to MSc award in Biology

Course period	From October 1st till February 1st, 1 semester (16 weeks)
Study credits	4 ECTS credits
Duration	144 hours
Language of instruction	English
Academic requirements	– BSc degree in Biology, Physics, Biophysics, Chemistry , Biochemistry, Environmental Sciences or equivalent (transcript of records), – good command of English (certificate or other official document)

Course Description

«Optimization and Data Analysis in Biology» is an extensive course, which is designed to advance a student's ability to interpret experimental biological data and to construct primary mathematical descriptions of the phenomenon under investigation. It provides with wide comprehension of statistical nature underlying biological data together with competence to operate basic mathematical models in biology.

The most important part of this course is stating and verification of statistically consistent hypotheses in biology. Students will differentiate between several possible explanations of a given phenomenon from an experiment to find the most appropriate one.

The "optimization" part of the course includes the inverse problem solution – determination of kinetic parameters of a process from the available experimental data and their mathematical description. Such an approach allows to evaluate characteristic velocities of processes in the experiment. The course is designed to cover a wide range of possible areas of biology and medicine. Completion of this course will help future specialists to construct and develop the explanation of observed phenomena.

Special Features of the Course

1. The course is friendly for the biologists

It does not require a deep knowledge in math and statistics. At the end of this course a student as a biologist-experimenter is expected to be able to perform the basic analysis of his/her results.

2. Provides a basic non-trivial analysis of the results

A standard set of parameters that experimenters use in the analysis is quite narrow. It usually includes construction of error bars, calculation of p-values and drawing an approximation line. The course provides students with more powerful methods in addition to this set, which will yield the deeper understanding of the process under investigation.

3. Flexible to the software

Students who are already familiar with such mathematical packages as MatLab, SAS, R etc. can use them during the course, rather to use SciLab. The course provides students with clear and basic schemes that could be easily converted into familiar code.

Course Aims

- To help students to systematize and repeat previously covered material of basic statistics, which is an important part of the course.
- To assist students to plan an experiment in the field of their interest.
- To provide students with understanding of hypotheses analysis.
- To introduce the basic principles of statistical and mathematical processing of experimental data.
- To give introduction of optimization methods and their application in biological experiments.

Course Objectives

- To ensure that students are familiar with basic statistical distributions.
- To give students optimal planning methods for experimental design.
- To give students an appreciation of testing algorithms and estimation criteria for hypothesis testing.
- To provide students with the basic routines of statistical and mathematical model construction, through data simulation and usage of ordinary differential equations systems.
- To equip students with knowledge and understanding of the main principles and parameters used for estimation of a model.

Learning Outcomes of the Course

By the end of the course, students will be able to:

- select an optimal set of experimental conditions to obtain maximum of meaningful information,
- compare experimental results derived in different conditions,

- construct mathematical and/or statistical representation of the observed process from experimental data,
- determine the main characteristics of experimental processes.

Course Outline

Week	Lectures	Practice session / Assignments	Hours[1]
1-4	Biological Data Mining	– Scientific method in biology with SciLab – SciLab basics – Experimental planning. How to collect the data – Home assignment No 1	2 4 2
5-7	Distributions in Biology	– Random events and distributions – Theoretical and experimental distributions – Distributions for hypotheses testing – Home assignment No 2	2 2 2
8-11	Comparison of samples	– Hypotheses in biology – Hypotheses testing – Interval estimations – Confidence intervals calculations – Home assignment No 3	2 2 2 2
12-15	Mathematical description of biological processes	– Experimental data approximation – Correlation, dependence and regression – Mathematical modelling in biology – Simple ODE mathematical models – Optimization techniques – Optimization quality and models comparison	2 2 2 2 2 2
16	Final exam		4

[1] Hours designed for Classroom sessions, Web-sessions, Home Assignments etc.

Lecturer and Contact Information

Andrey SHUVAEV

Ph.D., Associated Professor at School of Fundamental Biology and Biotechnology, Siberian Federal University

(room 13-08) 79, Svobodny prospect, Krasnoyarsk, Russia

Tel: +7 391 206 2072, AShuvaev@sfu-kras.ru

Assessment

- Problems can be solved with any suitable package: SciLab, Matlab, SAS, R etc., but only SciLab scripts are discussed during the course.
- Grade policy for both home assignments and the final exam is:
- A (excellent work) 91–100 points
- B (above average work) 81–90 points
- C (average work) 71–80 points
- D (below average work) 50–70 points
- F (failed work) < 50 points
- The final exam is an individual problem in the form of a simulated experimental dataset. Students should be able to:

  • compare the given sets by the proposition of the hypothesis and performing the suitable test for its verification (30 points maximum),

  • regression of the data with the predefined theoretical curves with calculation of the confidence intervals of the fitted function (20 points maximum),

  • statistical or ODEs model with optimization routine (40 point maximum)..

Attendance Policy

Students are expected to attend classes regularly. In case of missing in-lab activity a student should perform an additional exercise (will be given by the lecturer) within one week.

Every topic has a home assignment work that should be done in written form (except several questions in the first assignment). The report on the assignment should be submitted before the lecture within 5 days from the moment students received a list of problems. The final mark will be made by the same grade policy as for a final exam.

Lectures and seminars marked by gray color in the schedule table on page 6 are obligatory.

Web page of the course

The webpage of the course *Optimization and Data Analysis in Biology* is available through E-learning SibFU web site: www.e.sfu-kras.ru . You must be logged in to access this course. Course materials and required reading materials are available at the course web-page.

Core reading

The main book for this course (besides the Course Book) is «A Modern Introduction to Probability and Statistics» (ST) by Frederik Michel Dekking et al. It contains all information that is required for study, but in a more extensive manner than the Course Book. It will help students to reach a deeper understanding of methods and applications of data analysis together with optimization.

A book «Resampling Methods» by Phillip I. Good (RM) is recommended for studying the basic resampling techniques that will be used during the course.

Some of the course topics include practical tasks in Excel. Although this analysis is very basic, a book «Excel Data Analysis» (EDA) by Hector Guerrero can provide students with additional information.

The Course includes a lot of regression procedures that will be made mostly by least squares technique. Therefore «Least Squares in Data Analysis» (LS) by John Wolberg will be widely used.

As soon as the most of analysis and optimization techniques will be made in SciLab, an extensive open library must be used to reach the required ability in using of SciLab. Those students who are familiar with other packages are allowed to do the analysis using it. For R users the John Chambers «R-Software for Data Analysis» book might be useful. For SAS user the Mervyn G. Marasinghe, William J. Kennedy «SAS for Data Analysis» could be used.