MCB111 Mathematics in Biology (Fall 2024)

http://mcb111.org
Mon/Wed/Fri 10:30-11:45

Instructors: Dr. Elena Rivas elenarivas@fas.harvard.edu

Description

The divide between experimental biology and computational/quantitative biology is getting thinner and thinner, and (hopefully) will disappear in the next few year. This course is meant for biologists who want to learn mathematical principles relevant to current biological research as well as for mathematicians who want to find biological applications to mathematical principles. About half of the course covers topics on information theory, inference, statistics, and probabilistic modeling. The second half of the course covers dynamical systems in biology, including random walks, feedback control, and molecular population dynamics.
Each week-long unit is devoted to one specific topic, and is based in one or more scientific papers selected from the recent biological literature. Each unit includes a set of lectures (available online), a practical session, and a homework. The practical session follows a flipped-class model in which students work in the classroom implementing the methods described in the lecture.

Objectives

I will show you how to be critical with your data; and how to run the right control experiments. This course covers some of the mathematical tools to do that. I will use plain language as much as possible without losing mathematical rigor. I want to get across to you that you should not use statistical tests for which you do not understand the assumptions (and they all have them!), nor treat math as a black box. And if your calculus fails you and cannot find an analytic expression, then you can solve a problem numerically.
I also would like to show you that doing simple scripting is easy, and that in doing so, you can stop making assumptions about how simple models work, and instead see how they work. The course has an important emphasis on computational work. Computer literacy is fundamental for experimental biologists. You are expected to know some basics of one computer language of choice (Matlab, Python, Perl, or others), but you are not expected to be an expert on any one of them. By the end of the course, hopefully you will be confident launching into your own computational approaches to data collection, data analysis, and model testing.
There. Let's get started!

Syllabus

Week	Dates	Topic
W00	04/06 Sep	**Introduction to coding and probability distributions** *We will learn to sample from any probability distribution as an intro to coding.*
W01	09-13 Sep	**Information Theory.** *What is information content, and why it's good to assume ignorance!*
W02	16-20 Sep	**Probability and Inference.** *Using the data to reverse engineer your system.*
W03	23-27 Sep	**Model comparison - Hypothesis testing.** *Which model is best supported by my data?*
W04	30-04 Sept-Oct	**Significance: The Student's t-test and p-values.** *Behind a classical statistical test, and how to use null models.*
W05	07-11 Oct	**Maximum likelihood and least-squares.** *A principled derivation of linear regression.*
W06	16/18 Oct	**Probabilistic Models and Inference.** *A Hidden Markov Model to assign ancestry to chromosomal segments.*
W07	21-25 Oct	**Probabilistic Models and Expectation Maximization.** *How to train your model.*
W08	28-01 Oct-Nov	**Neural Networks. Learning as Inference.** *Where we will design, train and make inferences with a simple neural network.*
W09	04-08 Nov	**Random Walks in Biology.** *Diffusion: from the microscopic to the macroscopic.*
W10	11-15 Nov	**Molecular Population Dynamics as a Markov Process.** *mRNA birth-and-death (synthesis/degradation and steady state)*
W11	18-22 Nov	**Feedback Control in Biological Interactions.** *We will study several cases of gene regulation by positive and negative feedback loops.*
W12	25/02 Nov-Dec	**Gene switches.** *Synthetically designed E. coli switches and circadian rhythms in Drosophila.*
W13	04 Dec	**Pattern formation and Turing Instability.** *Making cool patterns by diffusion and reaction.*

Expectations

I expect attendance and completion of homework.

You can work with other students when completing your homework, but you have to present your own materials and code, showing the logic behind the coding statements.
As for using ChatGPT and other generative AI tools to produce your homework, these are the guidelines:

- You can look at AI generated materials at any time during the completion of your work, but you cannot present AI generated material as your own.

- These materials should be considered as a tool, sometimes useful for you to learn a technique or to see code implementing it. But be aware that these contents (unlike those in published books with authors responsible for them), are not produced with the expectation that they are correct or accurate.

- Ultimately, you are responsible for all work that you present as yours, and has to be writen in your own words.

Logistics

There will be one homework per week, and one final exam (open book). Grade will be based on homework (60%), final exam (30%), and participation (10%). Participation includes: attendance to class, participation in Studtent Hours or other forums such as piazza or slack.

Accademic Integrity

You are encouraged to disccus your homework with your classmates, but the work you present has to be your own and be written in your own words. If someone helped you with a particular aspect, add a comment in your code explaining who helped you and what is the contribution, and you want to demonstrate that you understand it. This also applies to any matierials taken from the internet.

Materials

The following materials are for reference. You are not expected to get them all, not even one of them. The class notes should contain all materials covered in this course; and specific reading materials will be proposed with each topic. I will try to have some of those books for you in reserve in the library so you can look at them. One of the books (MacKay's–one of the most insightful books ever, but not super easy to read) is free online.

For the first half of the course, I will use materials from these books:

1. "Data Analysis. A Bayesian Tutorial", D. S. Sivia J. Skilling, Oxford University Press, 2005.
2. "Information Theory, Inference, and Learning Algorithms", D. J. C. MacKay, Cambridge University Press, 2003. (available free online at http://www.inference.phy.cam.ac.uk/mackay/itila/book.html).
3. "Introduction to Probability", J. K. Blitzstein and J. Hwang, CRC Press, 2014.
4. "The Cartoon Guide to Statistics", L. Gonick and W. Smith, Harper Perennial, 1993.
5. "Statistical Distributions", 3rd edition, M Evans, N. Hastings, and B. Peacock, Wiley Interscience, 2000.
6. "Probability Theory. The Logic of Science", E. T. Jaynes, Cambridge University Press, 2003.

For the second half of the course, I will use materials mostly from these books:

1. "Physical Models of Living Systems", P. Nelson, W. H. Freeman and Company, 2015.
2. "Random Walks in Biology", H. C. Berg, Princeton University Press, 1993.

These days any mathematical results can be found online. Still, if you want an old fashion approach, I have always found this Schaum manual very useful,

1. “Mathematical handbook of formulas and tables”, 4th edition, M. R. Spiegel and S. Lipschutz and J. Liu, McGraw-Hill, 2013.

It is cheap, and it includes most of the mathematical expressions I need on a day-to-day basis for integrals, series, special functions, etc. And you can also find the pdf of the whole book online.