

MCB111 Mathematics in Biology (Spring 2018)

<http://mcb111.org>

Harvard Biolabs 1080, Mon/Wed 10-11:30

Instructors: Dr. Elena Rivas elenarivas@fas.harvard.edu
Pu Zheng pu_zheng@g.harvard.edu

Description

The barrier between experimental biology and computational/quantitative biology is getting thinner and thinner, and (hopefully) will disappear in the next few years. 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
W01	22-24 Jan	Introduction to Information Theory. <i>What is information content, and why it's good to assume ignorance!</i>
W02	29-31 Jan	Probability and Inference. <i>Using the data to reverse engineer your system.</i>
W03	05-07 Feb	Model comparison - Hypothesis testing. <i>Which model is best supported by my data?</i>
W04	12-14 Feb	Significance: The Student's t-test and p-values. <i>Behind a classical statistical test, and how to use null models.</i>
W05	21 Feb	Maximum likelihood and least-squares. <i>A principled derivation of linear regression.</i>
W06	26-28 Feb	Probabilistic Models and Inference. <i>A Hidden Markov Model to assign ancestry to chromosomal segments.</i>
W07	05-07 Mar	Probabilistic Models and Expectation Maximization. <i>How to train your model.</i>
	12-16 Mar	spring break
W08	19-21 Mar	Principal Component Analysis and Clustering. <i>When data needs to be grouped and you do not know how.</i>
W09	26-28 Mar	Random Walks in Biology. <i>Diffusion: from the microscopic to the macroscopic.</i>
W10	02-04 Apr	Molecular Population Dynamics as a Markov Process. <i>mRNA birth-and-death (synthesis/degradation and steady state)</i>
W11	09-11 Apr	Feedback Control in Biological Interactions. <i>We will study several cases of gene regulation by positive and negative feedback loops.</i>
W12	16-18 Apr	Gene switches. <i>Synthetically designed <i>E. coli</i> switches and circadian rhythms in <i>Drosophila</i>.</i>
W13	23-25 Apr	Pattern formation and Turing Instability. <i>Making cool patterns by diffusion and reaction.</i>

Expectations

I expect attendance and completion of homework. I will do my best to encourage participation in class. You must do work on your own.

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%).

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.