Housekeeping

BIOF2014

Schedule

Tuesdays & Thursdays 4:30 pm - 6:20 pm

Locations: MB217 or CPD-3.41

Motivations

I have had little formal training in math at the university level (1 year calculus + 0.5 year linear algebra) and no statistics.

In graduate school, I wanted to take statistics but the course instructor did not grant permission.

Learning mathematical statistics by yourself is very difficult.

The bioinformatics curriculum needed a new statistics course, and I could have just declined to teach this course.

Statistical modelling is *very* useful in bioinformatics, but many bioinformaticists don't have a solid foundation in statistics and statistics can take a long time to learn.

My goal: Provide a solid foundation for you to develop and use statistical models effectively.

Comparison

course	theory	calculation	derivation	programming
STAT2601	extensive	extensive	extensive	none
BIOF2014	extensive	minimal	extensive	extensive

Course philosophy

We will cover the most useful mathematical concepts that are useful for statistical modeling of real-world data.
Mathematics is how we communicate with one another about our models.
Programming is how we communicate with machines to do calculations.

Text books

Statistical Inference for Everyone. Brian Blais. 2014. https://open.umn.edu/opentextbooks/formats/619

Statistical Inference. George Casella, Roger L. Berger. 2002.

Assessments

type	weight	derivation	proof	programming
12 quizzes	10%	Y	?	N
8 assignments	60%	Y	Y	Y
1 exam	30%	Y	Y	N

Quizzes are your opportunity to practice for the exam.

Exam questions are similar to derivation and proof questions in the quizzes and the assignments.

You will be permitted to bring one double-sided A4 sheet of *hand-written* notes to the exam. No calculator, computer, phone, or other electronic devices are allowed in the exam.

Re-weighting

If your exam grade is higher than your overall grade for the quizzes (or assignments), a weight of 10% will be shifted from the quizzes (or assignments) to the exam.

scenario	quizzes	assignments	exam
default	10%	60%	30%
assignments $\geq$ exam $>$ quizzes	0%	60%	40%
quizzes $\geq$ exam $>$ assignments	10%	50%	40%
exam $>$ max(quizzes, assignments)	0%	50%	50%

Lecture recording

Probably not feasible.

ChatGPT

As per university policy, instructors should make assessments 'ChatGPT-proof'.

ChatGPT is excellent at writing meaningless fluff, and it is mediocre at logic. It is often overfitted to the questions in the training set.

ChatGPT can be useful for correcting grammar, improving writing, and perhaps as a second opinion.

ChatGPT does poorly (0-60%) on my questions for other courses.

ChatGPT answers are often vague. Their derivations seem helpful on the surface but lack mathematical rigor and contain hallucinated/magical steps.

Marking

Lecturers' jobs are to teach.

Professors' jobs are to win grants and produce research outputs.

No one wants to do any marking.

Traditionally, exams are worth 50-100% of the final grade.

Grades *almost surely* (i.e. with probability $= 1$) will stay unchanged after remarking, except in cases of adding error.

You are always welcome to ask questions about the *course content* or related topics!

Marking

Why can't we just give everyone an A+?

What are the consequences when competency and reward are de-coupled?

Marking scheme

Math is as objective as it gets.

Getting the final answer right is *not* sufficient for full mark.

Rigor and clarity are just as important as the final answer.

All questions must be answered with *mathematical rigor*.

Correct process $\Rightarrow$ correct results (fact)

Incorrect results $\Rightarrow$ incorrect process (contrapositive)

Correct results $\Rightarrow$ correct process (converse error)

Marking scheme

Derivation and programming questions

Process	Result	Mark
Correct	Correct	100%
Correct but incomplete	Null	10-90%
Mostly correct with errors	Correct	40-80%
Mostly correct with errors	Incorrect	20-60%
Incorrect	Correct	0-20%
Incorrect	Incorrect	0%

Marking scheme

Proofs

Process	Result	Mark
Correct and well-justified	Given	100%
Correct and well-justified but incomplete	Given	10-90%
Correct without sufficient justification	Given	10-60%
Correct with wrong justification	Given	0-30%
Incorrect	Given	0%

Marks for humility

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." – Mark Twain

For each assignment question, you can get 1 mark for writing "I don't know", provided that all the following conditions hold:

a. "I don't know" is the entirety of your answer
b. the question is worth more than 1 mark
c. the question is *not* a bonus question
d. your assignment is submitted on time

The same applies to the exam.

Late submissions

The penalty for late submission is 10% per day.

Quizzes or assignments will *not* be graded after any answer is discussed or released.

Answers to quizzes are released at 8 a.m. 7 days after the quiz is open.

Notations

Learning and using the correct mathematical notation are critical!

Notations are invented to make expressions less ambiguous and more concise!

In the bioinformatics field, method descriptions are atrocious because few people learned the proper notations, which make communications very difficult.