MATHEMATICS J-23C, Winter 2020-2021
Mathematics for Computation, Statistics, and Data Science

Last revised: December 8, 2020
Course Website: `https://canvas.harvard.edu/courses/79128`

Author of lectures and problems:

- Paul Bamberg `bamberg@tiac.net`

Head Teaching Assisant:

- Michael Liotti `mil727@g.harvard.edu`

Course Staff:

- Amy Benedetto `amybenedetto@college.harvard.edu`
- Rishabh Ghosh `rig685@g.harvard.edu`
- Raquel Ortiz `rao474@g.harvard.edu`
- Jean Sebastien Paul `jep577@g.harvard.edu`
- Vikram Singh `vikramsingh@college.harvard.edu`
- Chris Wirth `cwirth@college.harvard.edu`
- Alex Wong `alex.wong2015@gmail.com`
- Lara Zeng `lzeng@college.harvard.edu`

Sections (Tentative):

- Thursday, 3-5 p.m. led by Amy and Chris
- Thursday, 7:30-9:30 p.m., led by Raquel, Vikram, and Alex
- Friday, 7:30-9:30 p.m., led by Michael and Lara
- Saturday, 9-11 a.m., led by Rishabh and Jean Sebastien

Goals: Math 23c is an alternative second half of a moderately rigorous course in linear algebra, real analysis and multivariable calculus, designed for students who are serious about mathematics and interested in being able to prove the theorems that they use, but who are as much concerned about the application of mathematics in fields like computer science, statistics, data science, and economics as about "pure mathematics" for its own sake.

The winter-session variant of this course, using recorded lecture videos from the first offering of the course delivered in the spring of 2018, will cover the first five weeks of Math 23c as it was taught in its first two iterations. This material, covering topics that appear exclusively in Math 23c, spans the foundations of mathematical logic through probability over uncountably infinite sample spaces. This should largely reconcile the differences in coverage between Math 23b and Math 23c in a way that wouldn't be possible in a single term due to time constraints. This is the second offering of the J-term variant of the course. In the 2019 offering, all 19 students who completed the course and attempted the exam passed, many of them earning perfect or nearly perfect scores.

Math 23c meets the goals of the new Quantitative Reasoning with Data requirement for General Education. The legislation that established this requirement states:

"This requirement ensures that students reach a level of quantitative facility involving mathematical, statistical, and computational methods that will enable them to think critically about data as it is employed in fields of inquiry across the FAS and SEAS."

Math 23c is one of a handful of courses that cover "mathematical, statistical, and computational methods" in detail, and Math J-23c probably does more data analysis in five weeks than many courses that now meet the QRD requirement cover in an entire term.

Prerequisites: You should be skillful with sets, single-variable calculus and infinite series and have some experience with proofs. Linear algebra and multivariable calculus will be needed later in Math 23c, but not in the first five weeks.

Math 23a is optimal. Math 21a, plus a BC Calculus course that did a good job with sequences and series, should also provide an adequate background, but you may have to play catchup with proofs. Even Math 1b may suffice. There are lots of videos available on the Math 23a Web site. Students who plan to take 23c without 23a can use them to fill in gaps in their background.

Although everyone will be required to write simple R scripts, no previous programming experience is required. You just need to be an experienced computer user.

Classes

Math J-23c is using videos of the lectures from spring 2020. Classes, two hours per week, consist of a seminar (30-40 minutes) where students present key topics and a problem-solving workshop (80-90 minutes). These sessions will take place at the end of the week in which you are expected to have watched the recorded lectures. Members of the class will volunteer to present one of three seminar topics, which go over key definitions and proofs The second hour will be devoted to problem solving in small groups, wherein each group will be tasked with completing at least one part of each of the four numbered problems. Some of the problems will require the use of R. Section instructors will take attendance at the start of each section for the purpose of collecting data on the course, and though

this will not ultimately factor into your grade, it is not advisable to miss section.

Over the winter term, these sections will take place over Zoom web conferencing. You will need a camera and a microphone in order to present proofs to and interact with your classmates and section instructor. There are several clever strategies that students have come up with in past years for this:

- logging an iPad or tablet into Zoom and using either an application (such as OneNote) or the Zoom whiteboard;
- logging a phone into Zoom and pointing it at a piece of paper, perhaps with the aid of a phone clamp;
- writing or typing the proof in advance and scanning it in, using the 'Share Screen' feature on Zoom to talk through the finer details;
- creating slides using PowerPoint, Beamer, or another, similar application.

You are strongly advised to test that your machine is working appropriately prior to the start of sections, especially if you are only accustomed to on-campus sections. You can sign up for Zoom at `https://zoom.us/`.

Sections are strictly optional for all students, extension or FAS. You may decide to watch the lecture videos on your own, read the solutions, and attend office hours in preparation for the two exams. Unlike in a full-term variant of the course, your grade will be computed identically to the students who are sectioned. Even if you are only able to attend a fraction of section meetings, we advise you to join a section as a way of keeping up with the material. You are free to attend another section in any week in which you have a time conflict, provided that you have the permission of the section instructor(s).

Paul delivered new lectures for Math 23c last spring, which presented mathematical content alongside R markdown. You may watch either these lectures or the previously recorded lectures from the spring of 2018, but if you opt for the latter, you will likely need to supplement with Paul's narrated R script videos. Though the spring version of this class does not feature a seminar, the January term will, the intention of which is to best prepare you for the final exam, which will include several key proofs from lecture.

For each lecture, there is also a scan of Paul's lecture notes on the Web site. Consensus is that you do will do best to take your own notes and use these only as backup. Many students report that it is efficient to play the easy parts of the lecture at double speed, and then to slow down and perhaps even run twice though difficult proofs or intricate examples.

R Script videos, of which there are typically two per week totaling about 40 minutes, will introduce techniques for data visualization and analysis and show applications of mathematical ideas that would be tedious to do without computer assistance. These are important background for workshop problems.

No prior experience with R is assumed. Students who worked with R in Math 23a will have an easier time, but the set of R skills to be covered overlaps little with what was done in Math 23a.

Proof Logging: Unlike in Math 23a, proof logging will **not** run in Math J-23c nor in the spring variant of Math 23c. Even with that said, each week's three seminar topics may appear on the exam, so students may find it useful to get together and present proofs to each other. Course assistants may choose to organize sessions dedicated to the course's proofs.

Exams:

Math J-23c will have two exams. The first will be a short midterm exam, written by Michael, at the end of the third week intended to keep people on track with the material. It will comprise 15% of your overall score. If your final exam grade is materially higher than your midterm exam score, it will take the place of your weighted score.

The final exam, written by Paul, will be given at the end of the fifth week on the material covered over this session, encompassing the first five weeks of the full-semester Math 23c/E-23c course. The final exam will be available, starting the weekend prior to the start of spring classes, over a period of a few days. You will need to find an informal proctor (roommates or family members are OK) who will read the Harvard College Honor Code and sign a statement that you completed the quiz under closed-book conditions, with computers and cell phones turned off. You must complete the quiz at a single sitting during the time window in which it is available, but there is otherwise no time limit. Two hours should be sufficient.

The two exams will be the only assessment, formal or informal, of Math J-23c. You will not graded on section attendance or participation, nor will you be required to submit problem sets. The exams will include questions that resemble the ones done in workshops, and some seminar topics will appear verbatim. There may be other short proofs similar to ones that were done in lecture.

If you final grade is at least 75% and you opt to take E-23b, your exam score will be averaged in if the effect is to improve your grade, and you will satisfy the requirement of extra work for graduate credit in E-23b. If you opt to take Math 23c/E-23c, you may opt to have your score included in the calculation of your spring grade as an extra 40-point quiz grade. Given that there will only be two quizzes and a term project in Math 23c replacing the final exam, you may find that this will be of immense benefit to your grade. Furthermore, you will be extremely well-prepared for the first five weeks of the course and may even have completed the assignments before the course starts. You are nonetheless advised to attend or watch Paul's spring lectures, which will differ from past year's recorded lectures.

Textbook:

Vector Calculus, Linear Algebra, and Differential Forms, Hubbard and Hubbard, fifth edition, Matrix Editions, 2015.

This book is in stock at the Coop, or you can order it at a lower price from the publisher's Web site at `http://matrixeditions.com`.

This book will see little use, but everyone who takes Math 23 a,b, or c needs to own it anyway

Other useful books:

Ross, Elementary Analysis: The Theory of Calculus, 2nd Edition, 2013.

This will be the primary reference for Week 4. It was the real analysis text for Math 23a and will be a useful reference for those with no formal background in real analysis. It is available electronically through the Harvard library system (use HOLLIS and search for the author and title).

Chihara and Hesterberg, Mathematical Statistics with Resampling and R

This book was the inspiration for many of the data science examples in the R scripts, and there will be occasional references to it in the scripts.

It is available electronically through the Harvard library system (use HOLLIS and search for the author and title). Used copies can be sometimes be found on amazon.com for as little as $49.

R and RStudio:

R is free, open-source software. Instructions for download and installation are on the Web site. You start to use R at the first workshop, so install it right away, preferably on a laptop computer that you can bring to the workshop.

On the course Website are a set of R scripts, with accompanying YouTube videos, that explain how to do almost every topic in the course by using R.

There will be workshop problems anf homework problems each week that require creation of R scripts.

Collaboration policy:

Since problem sets will not be handed in or graded, it is a very good idea to discuss problems with your classmates.

If you take Math 23c in the spring term, complete your problem sets early and hand them in before the official deadline.

Given that some of the students in this course will be taking Math 23c or E-23c for credit, we will not post solutions to the problem sets. There will be solutions posted for the seminar and workshop problems. These should not be posted to the internet nor shared with students who are not taking Math J-23c. Students who take Math 23c or Math E-23c in the spring will have access to the official, complete set of solutions upon completion of each week's problem sets. However, students in Math J-23c are encouraged to discuss any of the problems with one another or with the course assistants in preparation for the exam. This collaboration will be facilitated by publicly-available discussion boards, sections, and office hours. Be sure, however, that you thoroughly document all of those with whom you have collaborated during this period, in case you decide to submit those particular problem sets for credit in the spring term. In other words, you are bound by the same academic integrity standards over the J-term as you would

over the regular semester.

Schedule:

Month	Thursday date	Topic
December	⋆	Logic, using boolean algebra and finite fields
January	31	Foundations of probability for finite sample spaces
January	7	Probability for countably infinite sample spaces
January	9-11	EXAM on weeks 1-3
January	14	The Riemann integral and its generalizations
January	21 †	Probability for uncountably infinite sample spaces
January	23-25	EXAM on weeks 1-5

⋆ Due to the holiday, no sections will take place on Thursday, December 24th or Friday, December 25th. Instead, week 1 sections will run on Monday and Tuesday of this week, December 21st and 22nd, respectively.

† To avoid conflicts with online shopping week, which begins in the week of January 18th, optional make-up sections for week 5 will be run in the evening of Monday and Tuesday, January 18th and 19th, respectively.