Math 340 / 640
Fall 2023 (4 credits)
Course Syllabus

Adam Kapelner, PhD.

Queens College, City University of New York

document last updated Tuesday 22nd August, 2023 9:24pm

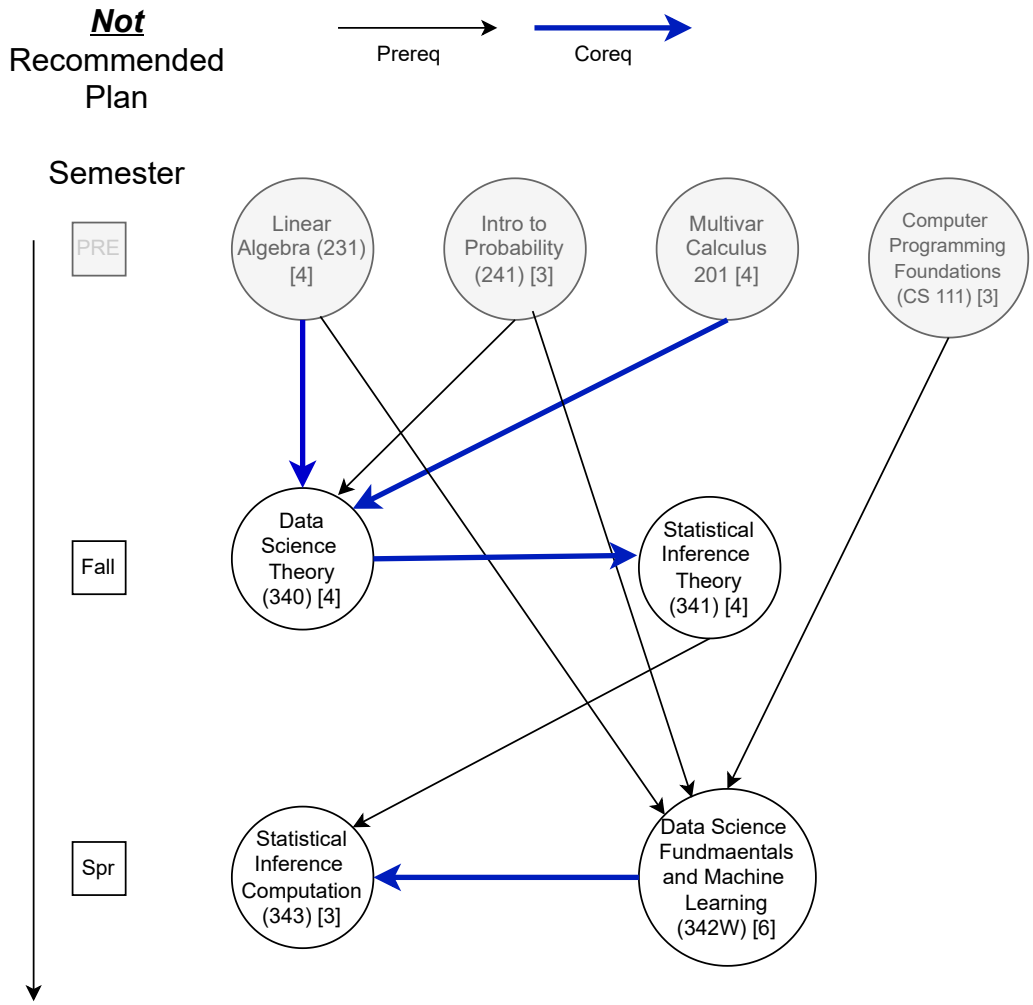
Course Homepage	`https://github.com/kapelner/QC_Math_340_Fall_2023`
Slack Homepage	`https://QCMath340Fall2023.slack.com/`
Contact	kapelner@qc.cuny.edu
Office	604 Kiely Hall
Lecture Time and Loc	e.g. Tues and Thurs 3:10 – 5PM / KY283
Instructor Office Hours and Loc	see course homepage

Course Overview

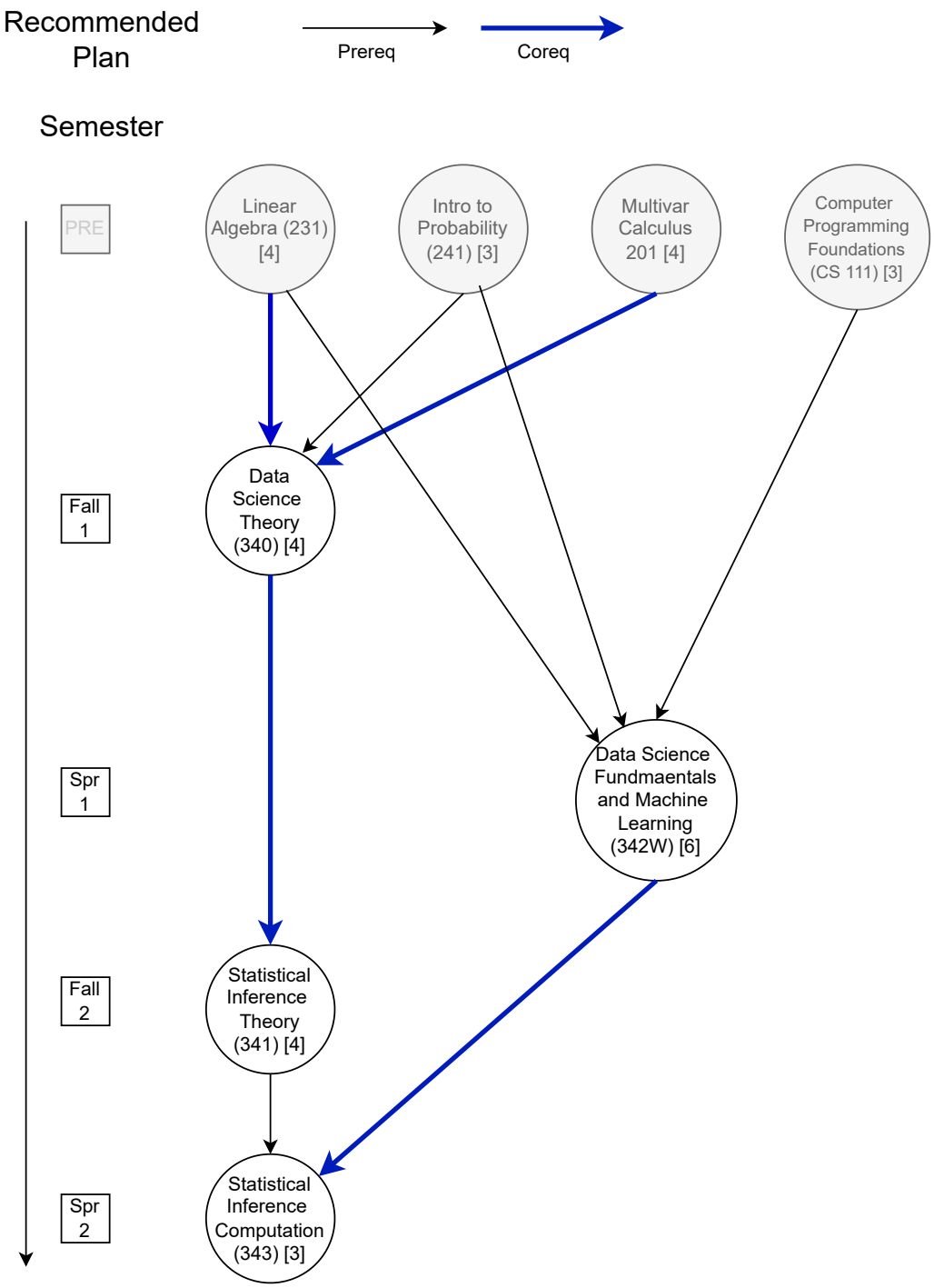
MATH 340 / 640. Probability Theory for Data Science and Statistics. 4 hr.; 4 cr. Prereq.: 241. Coreq.: MATH 201 and 231. Convolutions, multivariate transformation of variables, the poisson process, beta, logistic, laplace, Weibull distributions, characteristic functions, central limit theorem, Cochran's Theorem, the multivariate normal, chi-squared, T, F, Cauchy distributions, vector random variables, covariance matrices, multinomial, multivariate normal, order statistics, law of large numbers, famous inequalities, convergence in distribution and probability, laws of large numbers, Slutsky's Theorem, optimization including stochastic gradient descent, Markov chains, Gibbs sampling. Probability computation using modern software. Special topics. Not open to students who are taking or who have received credit for MATH 340. Students cannot receive credit for more both: MATH 340, and 640. Fall, Spring.

The Four Data Science Core Classes

This course is one of the four data science core courses but it does not cover any typical "data science" topics. Instead, it is designed to provide theoretical skills for Math 341 and 343. Math 340 and 342W are designed to be standalone courses and the other two courses rely on topics covered therein. Thus there is an order the classes need to be taken. Below are two plans, the first is over two semesters and hence it is not *not recommended* as it will be a *very* heavy workload. The second is over four semesters and it is the recommended plan as I believe it will allow students to absorb the material more effectively:



Examining the above, we note that MATH 340 can be taken as a standalone course as a higher-level math elective. It can be thought of as advanced probability following MATH 241. This course is also recommended for students who are considering an actuarial career as this course provides great practice for Exam P.


Recommended Plan
Prereq
Coreq
Semester
PRE
Fall 1
Spr 1
Fall 2
Spr 2
Linear Algebra (231) [4]
Intro to Probability (241) [3]
Multivar Calculus 201 [4]
Computer Programming Foundations (CS 111) [3]
Data Science Theory (340) [4]
Data Science Fundmaentals and Machine Learning (342W) [6]
Statistical Inference Theory (341) [4]
Statistical Inference Computation (343) [3]

Tentative Day-by-Day Schedule

Lectures and their topics with rough time estimates per topic are below:

Day 1 [10min] Review of discrete random variables (rvs), support; [10min] the Bernoulli rv, parameter space, degenerate rv, indicator functions; [10min] old-style probability mass functions (PMFs), new-style PMFs using the indicator function; [10min] cumulative distribution functions (CDFs), survival functions, [10min] vector rvs, joint mass functions (JMFs), independence, identical distributedness, iid; [25min] sum of two Bernoulli rvs, convolution operator, tree diagram for two rvs, conditional probability, marginal probability; [10min] Plotting the PMF of the convolution; [20min] discrete convolution formulas (general, independent rvs, iid rvs, old and new style), convolution support

Day 2 [5min] visualizing convolutions as passing PMFs "through each other"; [5min] defining combinatorial terms with indicator functions; [35min] Binomial rv with PMF derivation, Pascal's identity, Vandermonde's identity; [20min] Sequences of rv's, derivation of the geometric rv as a "waiting time" or "survival" distribution; [30min] convolution of geometric rvs, the negative binomial rv

Day 3 [25min] derivation of the Poisson rv, convolution of Poisson rvs; [30min] Strategy to compute $\mathbb{P}(X > Y)$ for two rvs X, Y, the sum reindexing trick, geometric series; [25min] multinomial rv, multichoose notation; [20min] proof that multinomial's marginals are Binomial

Day 4 [30min] Proof that conditional multinomials are multinomial, indicator functions for undefinedness; [25min] review of expectation, variance, standard deviation, covariance of two rvs, covariance rules; [5min] expectation of vector rvs; [15min] variance-covariance matrix (varcov); [5min] rules of expectation of vector rvs; [15min] rules of varcov and quadratic forms; [5min] Markowitz optimal portfolio theory application

Day 5 [25] expectation and varcov of Multinomial rv; [5min] Uniform discrete rv; [15min] The negative of a discrete rv; [30min] derivation of the Skellam rv as the difference of two poissons; [10min] Conditional-on-sum calculation; [30min] Derivation of exponential rv from geometric via the CDF, non-existence of PMF, review of continuous rvs and the PDF

Day 6 [35min] derivation of convolution formulas for continuous rvs; [35min] continuous uniform rv, convolution of two continuous uniform rvs; [25min] convolution of exponential rvs to derive the Erlang rv

Day 7 [30min] gamma function, incomplete gamma functions, regularized incomplete gamma function, gamma function properties and identities; [20min] CDF of Erlang, CDF of Poisson; [15min] Poisson Process; [10min] Gamma rv and Extended Negative Binomial rv; [15min] transformations of discrete rvs to derive PMFs

Day 8 Midterm I Review (Wednesday, Sept 27)

Day 9 Midterm I (Monday, Oct 2)

Day 10 [30min] transformations of continuous rvs to derive CDFs and PDFs; [10min] shift and scale PDF formulas for continuous rvs; [10min] derivation of the Logistic rv; [10min] logistic function and Logistic rv CDF; [25 min] applied project: use Logistic to model chess ratings; [35min] quantiles, percentiles and the median, quantile function for continous rvs

Day 11 [20min] derivation of the Pareto rv PDF, CDF and quantile function; [30min] applied project: use Pareto to model land ownership; [15min] derivation of the Laplace rv; [10min] error distributions; [15min] derivation of the Weibull rv, memorylessness and the Weibull modulus; [20min] applied project: use Weibull to model human lifespans

Day 12 [50min] order statistics, CDF/PDF of minimum and maximum, CDF of arbitrary order statistics; [15min] derivation of the PDF of arbitrary order statistics; [15min] order statistics of iid Uniform rvs to derive the Beta rv; [15min] kernels of PMFs/PDFs; [15min] proof that convolution of Gamma rvs is a Gamma rv

Day 13 [15min] the beta function, incomplete beta function, regularized beta function, the Gamma-Beta function identity, the CDF of the Beta rv; [35min] Multivariate transformations of rvs, Jacobian determinants; [10min] PDF of ratio of rv formulas; [10min] PDF of proportion of sum of rv formulas; [10min] proof that the proportion of sum of two Gamma rvs is a Beta rv; [10min] proof that the ratio of two Gamma rvs is a BetaPrime rv

Day 14 [35min] mixture and compound rvs; [10min] derivation that the Poisson rv compounded with a Gamma rv is a ExtendedNegativeBinomial rv; [15min] derivation that the Binomial rv compounded with a Beta rv is the BetaBinomial rv; [15min] review of complex numbers, proof of Euler's famous formula relating 0, 1, e, i and π; [20min] basics of Fourier transform / analysis and Fourier inverse transform / synthesis, L1 integrable functions, statement of the Fourier inversion theorem

Day 15 [20min] demo of Fourier analysis with singing head and chest voices and the same note on the electric piano; [40min] definition of characteristic functions (chfs) and properties 0, 1, ..., 8, definition of convergence in distribution for Levy's Continuity theorem; [5min] definition of moment generating functions (mgfs) and properties 0, 1, ..., 4; [10min] derivation of the chf for the Gamma rv and proof that two Gamma rvs convolved is a Gamma rv; [25min] Proof of the central limit theorem (CLT): derivation of the limiting characteristic function

Day 16 [25min] Proof of the central limit theorem (CLT): inverting the limiting characteristic function to derive the PDF of the standard Normal rv; [15min] derivation of the general Normal rv, its moments, its chf; [10min] proof that the convolution of two independent Normal rvs is a Normal rv; [5min] derivation of the LogNormal rv and an application in portfolio theory; [25min] derivation of the χ^2_n rv and its equivalence to a Gamma rv, derivation that scaled Gamma rvs are Gamma rvs, derivation of the χ_n rv

Day 17 [20min] derivation of the F rv as the ratio of scaled χ^2_n rvs; [20min] derivation of Student's T rv as the ratio of a Normal rv to a scaled χ rv; [10min] derivation of the standard and

general Cauchy rv; [5min] Proof that the expectation of the Cauchy rv doesn't exist; [10min] the physicist's derivation of the Cauchy rv; [15min] derivation that the scaled sample variance is a χ^2_n rv; [15min] quadratic forms embedded in χ^2_n; [15min] Cochran's theorem introduction

Day 18 Midterm II Review (Wednesday, Nov 1)

Day 19 Midterm II (Monday, Nov 6)

Day 20 [30min] Cochran's thm proof including simultaneous diagonalization from linear algebra; [15min] the independent of sample variance and sample average in the case of iid normal rvs; [5min] Derivation of that the t statistic is indeed distributed as Student's T rv; [10min] derivation of the standard MultivariateNormal rv; [20min] derivation of the general MultivariateNormal rv; [10min] multivariate chfs; [10min] chf of the MultivariateNormal rv; [10min] proof that the shifted and scaled MultivariateNormal rv is another MultivariateNormal rv; [10min] Mahalanobis distance

Day 21 [10min] derivation of multivariate chf property that finds conditional distributions; [55min] derivation of the Markov tail bound and its corrollaries including Chebyshev's tail bound and Chernoff's tail bound; [10min] tail bounds for the Exponential rv; [20min] derivation of the Cauchy-Schwartz inequality and proof that correlation is bounded between -1 and +1; [20min] Jensen's inequality

Day 22 [25min] convergence in distribution, proofs for PMF and CDF convergences; [10min] Convergence in probability to a constant; [15min] weak weak law of large numbers (WLLN); [40min] convergence in law, proof that convergence in law implies convergence in probability;

Day 23 [30min] Other useful inequalities; [20min] Continuous Mapping Theorem in one dimension; [20min] Continuous Mapping Theorem in multiple dimensions; [20min] Slutsky's theorem in one dimension

Day 24 [20min] Slutsky's theorem corrolaries e.g. sums, products; [20min] Other useful theorems for convergence in probability and distribution; [15min] law of iterated expectation; [15min] law of total variance for two rv's; [15min] general law of total variance for more than two rv's

Day 25 [45 min] theory of learning and estimation with concentration inequalities e.g. for the binomial parameter; [30min] introduction to linear programming; [35min] simplex method

Day 26 [30min] introduction to convex optimization; [40min] gradient descent; [30min] stochastic gradient descent

Day 27 [30min] Levenberg-Marquardt Algorithm (LMA); [30min] simulated annealing; [30min] evolutionary methods and particle swarm methods; [20min] Limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) method

Day 28 Final Review

This is more of a typical mathematics theory course than the rest of the data science series. But we will still attempt to keep our eye on developing ideas and concepts for helping to make decisions in the real world. Thus we may make limited use of computation using the `R` statistical language.

Prerequisites

MATH 241 (basic probability), 201 (multivariable calculus) and 231 (linear algebra) or equivalents. I expect a 241 class that covers more or less what I cover in 241. See the course homepage for links under "prerequisite review". The multivariable calculus and linear algebra we will use I will try to review in class.

Course Materials

Textbook: Introduction to Probability Theory by Hoel, Port & Stone. This book is out of print but you can buy it used on Amazon for ≈\$20, a reasonable price (as far as textbooks go). There is no excuse not to have this book. It is *required*. However, I will not ususally be teaching "from the book" — most of the material in the class comes from the lecture notes. The textbook is a way to get "another take" on the material. The textbook covers about only half of the material done in class (yes, sometimes we will be following the textbook page by page). For the other half, you will have to make use of other resources.

Computer Software: We will also be using `R` which is a free, open source statistical programming language and console. You can download it from: `http://cran.mirrors.hoobly.com/`. I do not expect you to do *any* programming. I will be giving you `R` code to run and expect you to interpret the results based on concepts explained during the course.

Calculator: You can use a TI-84, 85, 89 or any calculator which you wish. I strongly suggest you use Wolfram Alpha and its smartphone app.

The crosslisted "640" Section(s)

You are the students taking this course as part of a masters degree in mathematics. Thus, there will be extra homework problems for you and you will be graded on a separate curve.

The Use of Slack and Github as a Learning Management System

As the course homepage is updated (e.g. a new homework assignment is posted), you will hear about it in slack. You will also find the video recordings of lectures there. (If there are multiple sections of the class, only one section's lectures will be recorded). Each assignment will have its own channel. You can feel free to discuss things with your fellow students there. If you are asking me a question, you must do so in the `#discussions` channel for a general questions or the assignment-specific channel (e.g. `#HW03`) so other students can see the question and benefit from the answer. If you pm me, I will not answer and just ask you to move it to the

public channel. Do not be afraid to ask questions. There are many people who will have your same question!

Slack is a wildly successful company that recently got bought by Salesforce because businesses *really* use it. Pretend you are working at one of these businesses: no posting about random stuff; keep things professional!

We will not be using any features of github for learning management. Do not open "issues" on github!

Announcements

Course announcements will be made via slack in the `#general` channel (not on email). I am known to send a few slack messages per week on important issues. The Slack workspace `#general` channel is also connected to the course homepage via an integration. So every time I change the homepage (e.g. to release a new homework or upload notes or a video), you will get a notification.

I can't stress the following enough: if you are not on slack, you will miss all class announcements!!! Slack notifies you when there are messages. You may wish to mute all channels except for #general. See this article for how to do that.

Class Meetings

There are 28 scheduled meetings. Of these, 23 will be lectures, 2 will be midterm exams which are in class and 3 will be review periods during the meeting before the exams (see lecture schedule section above). The exam schedule is given on page 10. The last class of the semester will be rescheduled to be a review period that is conveniently before the final. We will discuss later in the semester.

Homework

There will be 6–9 theory homework assignments. Homeworks will be assigned and placed on the course homepage and will usually be due a week later in class. Homework will be **graded** out of 100 with extra credit getting scores possibly > 100. I will be doing the grading and will grade an *arbitrary subset of the assignment* which is determined after the homework is handed in.

Homework must be handed in by emailing it to me as a PDF. You must do one of two things:

- **Print out the homework and handwrite your answers in the alotted space for each question. Then scan your homework as a PDF. There are a ton of good photo-PDF apps for both iPhone and droid.**
- **Open the PDF on your device and use a PDF-editing program to electronically handwrite your answers and save the PDF.**

I will NOT accept homework that is not atop the original rendered homework PDF file. Remember to write your name. There are no regrades during this pandemic semester. Homework must be at maximum 5MB.

You are encouraged to seek help from me if you have questions. After class and office hours are good times. You are highly recommended to work with each other and help each other. You must, however, submit your own solutions, *with your own write-up* and in *your own words*. There can be no collaboration on the actual *writing*. Failure to comply will result in severe penalties. The university honor code is something I take seriously and I send people to the Dean every semester for violations.

Philosophy of Homework

Homework is the *most* important part of this course.[1] Success in Statistics and Mathematics courses comes from experience in working with and thinking about the concepts. It's kind of like weightlifting; you have to lift weights to build muscles. My job as an instructor is to provide assistance through your zone of proximal development. With me, you can grow more than you can alone. To this effect, homework problems are color coded green for easy, yellow for harder, red for challenging and purple for extra credit. You need to know how to do all the greens by yourself. If you've been to class and took notes, they are a joke. Yellows and reds: feel free to work with others. Only do extra credits if you have already finished the assignment. The "[Optional]" problems are for extra practice — highly recommended for exam study.

Time Spent on Homework

This is a three credit course. Thus, the amount of work outside of the 4hr in-class time per week is 6-9 hours. I will aim for 7.5hr of homework per week on average. However, doing the homework well is your sole responsibility since I will make sure that by doing the homework you will study and understand the concepts in the lectures and you won't have all that much to do when the exams roll around.

Late Assignment Policy

Late homework will be penalized 10 points per business day (Monday–Friday save holidays) for a maximum of five days. *Do not ask for extensions*; just hand in the homework late. After five days, **you can hand it in whenever you want** until *the last scheduled class meeting according to the official academic calendar*. As far as I know, this is one of the most lenient and flexible homework policies in college. I realize things come up. Do not abuse this policy; you will fall far, far behind.

LaTeX Homework Bonus Points

Part of good mathematics is its beautiful presentation. Thus, there will be a 1–5 point bonus added to your theory homework grade for typing up your homework using the LaTeX typeset-

[1] In one student's observation, I give a "mind-blowing homework" every week.

ting system based on the elegance of your presentation. The bonus is arbitrarily determined by me.

I recommend using overleaf to write up your homeworks (make sure you upload both the hw#.tex and the preamble.tex file). This has the advantage of (a) not having to install anything on your computer and thus not having to maintain your LaTeX installation (b) allowing easy collaboration with others (c) alway having a backup of your work since it's always on the cloud. If you insist to have LaTeX running on your computer, you can download it for Windows here and for MAC here. For editing and producing PDF's, I recommend TeXworks which can be downloaded here. Please use the LaTeX code provided on the course homepage for each homework assignment.

If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

Since this is extra credit, do not ask me for help in setting up your computer with LaTeX in class or in office hours. Also, **never share your LaTeX code with other students** — it is cheating and if you are found I will take it seriously.

Homework Extra Credit

There will be many extra credit questions sprinkled throughout the homeworks. They will be worth a variable number of points arbitrarily assigned based on my perceived difficulty of the exercise. Homework scores in the 140's are not unheard of. They are a good boost to your grade; I once had a student go from a B to and A- based on these bonuses.

Examinations

Examinations are solely based on homeworks (which are rooted in the lectures)! If you can do all the green and yellow problems on the homeworks, the exams should not present any challenge. I will *never* give you exam problems on concepts which you have not seen at home on one of the weekly homework assignments. There will be three exams and the schedule is below.

Exam Schedule

- Midterm examination I will be on [see course homepage] with the first review session on the class meeting prior
- Midterm examination II will be [see course homepage] with a review on the class meeting prior.
- The final examination will be [see course homepage] with a review TBA.

Exam Policies and Materials

I allow you to bring any calculator you wish but it cannot be your phone. The only other items allowed are pencil and eraser. I do not recommend using pen but it is allowed. Food is not allowed during exams but beverages are allowed.

I also allow "cheat sheets" on examinations. For both midterms, you are allowed to bring *one* 8.5" × 11" sheet of paper (front and back). Two sheets single sided are not allowed. On this paper you can write anything you would like which you believe will help you on the exam. For the final, you are allowed to bring *three* 8.5" × 11" sheet of paper (front and back). Six sheets single sided are not allowed. I will be handing back the cheat sheets so you can reuse your midterm cheat sheets for the final if you wish.

Cheating on Exams

If I catch you cheating, you can either take a zero on the exam, or you can roll the dice before the University Honor Council who may choose to suspend you. I have extremely elaborate means to detect cheating on my zoom exams. These means have been iterated upon and improved for three semesters. I suggest you don't try me.

Missing Exams

There are no make-up exams. If you miss the exam, you get a zero. If you are sick, I need documentation of your visit to a hospital or doctor. Expect me to call the doctor or hospital to verify the legitimacy of your note.

Special Services

If you are a student who takes exams at the special services center, I need to see your blue slip or receive email evidence one week before the exam to make proper arrangements with the center.

Class Participation

This portion of your grade is assessed based on your level of interaction during the course lectures e.g. asking and answering questions. Participation on slack also counts towards this total.

Grading and Grading Policy

Your course grade will be calculated based on the percentages as follows:

Homework	10%
Midterm Examination I	23%
Midterm Examination II	23%
Final Examination	39%
Class participation	5%

The semester is split into three periods :

(a) From the beginning until midterm I. Midterm I covers material during this time.

(b) From midterm I to midterm II. Midterm II covers material in this period only.

(c) From midterm II until the final. The final is cumulative and covers all course material.

Each of the periods is assessed evenly. Thus, each period must count the same towards your grade. Since there is 75% of the grade allotted to exams, there is 25% allotted to each period. Thus, the final is upweighted towards the material covered in the third period. In summary, the final will have 5/35 points $\approx$ 14% for the first period's material, 5/35 points $\approx$ 14% for the second period's material and 25/35 points $\approx$ 71% for the last period's material. A good strategy for the final is to just study the material after Midterm II and minimal studying for the previous material.

The Grade Distribution

As this is a small and advanced class, the class curve will be quite generous. Last year, it was approximately 40% A's and 40% B's. If you do your homework and demonstrate understanding on the exams, you should expect to be rewarded with an A or a B. C's are for those who "dropped out" somewhere mid-semester or who cannot demonstrate basic understanding. I have never given an F in this class. Don't give me a reason to change this tradition!

Checking your grade and class standing

You can always check your grades in real-time using `http://qc.gradesly.com`. You will enter in your QC ID number (or CUNYfirst email address). I will provide you with your password by email after the first assignment is completed.

Auditing

Auditors are welcome. They are encouraged to do all homework assignments. I will even grade them. Note that the university does not allow auditors to take examinations.