MATH 369 / 650 Fall 2020 (3 credits)
Course Syllabus

Adam Kapelner, Ph.D.

Queens College, City University of New York

document last updated Monday 14th December, 2020 2:55pm

Instructor	Professor Adam Kapelner
Contact	`kapelner@qc.cuny.edu` and at `#discussions` in our slack workspace
Time / Loc	Monday and Wednesday 5-6:15PM on zoom
My Office Hours / Loc	Monday and Wednesday 4:25-4:55PM on zoom
Course Homepage	https://github.com/kapelner/QC_Math_369_Fall_2020

Course Overview

MATH 369 is an introduction to the intermediate concepts of mathematical statistics, statistical inference and theory of experimentation. Main topics are listed below:

- Estimation
  - Data generating processes, parameters, estimators, estimates, samples, population
  - Point parameter estimators and estimates: loss, risk, bias-variance decomposition
  - Likelihood, log-likelihood, score function
  - Un/biased estimators and asymptotically unbiased estimators, consistent estimator
  - Method of Moments (MM), Maximum likelihood estimator (MLE)
  - Cramer-Rao Lower Bound, Fisher information, UMVUEs
  - Asymptotical normality and efficiency of MLE's, uni/multivariate delta methods
  - The Empirical CDF estimator
  - The non-parametric bootstrap

- Testing
  - Neyman-Pearson paradigm of Hypothesis testing
  - Left-sided/tailed, right-sided/tailed, two-sided/tailed tests
  - One proportion z-test, One sample z-test, one sample t-test, two sample z-test (un/equal variances), two sample t-test (un/equal variances), two-proportion z-test.
  - Size (α), Fisher's p-value and "statistical significance"
  - Statistical power and the power function
  - Effect sizes and "practical / clinical significance"
  - uniformly most powerful tests?, equivalence testing?
  - Wald, score and the generalized likelihood ratio testing procedures
  - Chi-squared tests for goodness of fit and independence
  - Kolmogorov-Smirnov's goodness of fit test and distribution equivalence test
  - Permutation Tests
  - Multiple testing comparison problem: familywise error control via the Bonferroni, Sidak and Simes procedures, false discovery rate control via the Benjamini-Hochberg procedure
  - The non-parametric bootstrap
- Experimentation and Causality
  - Casual inference vs statistical inference
  - Neyman-Rubin model, the counterfactual
  - Observational study, confounding variable, biased causal estimation
  - Experimentation, completely randomized experiment, unbiased causal estimation
- Model Selection
  - Naive model selection bias based on number of parameters (overfitting)
  - Model section criterions: AIC, AICC
  - Akaike weights / scores

Individual lectures' topics with time estimates are below:

Lec 1 [20min] Surveys, populations, sampling, sample size n, representativeness, SRS; [30min] definition of parameter, parameter space, inference, statistic, the three goals of statistical inference: point estimation, theory testing and confidence set construction.

Lec 2 [20min] Sampling with/without replacement, iid assumption, equivalence of populations and data generating processes (DGPs); [55min] estimators, point estimation metrics: biasedness and unbiasedness, loss functions (absolute, squared error, others), risk function, mean squared error (MSE), bias-variance decomposition of MSE, maximum risk for iid Bernoulli model

Lec 3 [35min] Introduction to hypothesis testing, the intellectual honesty of assuming the null hypothesis H_0, the theory you wish to prove H_a, right, left and two-sided tests of single parameters, test statistics; [30min] the Binomial exact test of one proportion in the iid Bernoulli DGP, retainment region (RET), rejection rejection, statistical significance; [10min] Type I error, α, Type II error, 2×2 confusion table of testing errors.

Lec 4 [10min] review of central limit theorem (CLT); [20min] employing the CLT to create the approximate one-proportion z-test; [20min] Fisher's p-value; [25min] Definition of power, assuming an H_a to compute the probability of Type II errors and power

Lec 5 [15min] The power function for a one-sided z-test; [15min] the iid normal DGP; [15min] exact one-sample z-test with known variance, its power function; [30min] the naive estimator for σ^2, proof of its biasedness, definition of asymptotic unbiasedness, Bessel's correction and the unbiased sample variance estimator S^2

Lec 6 [20min] motivation of the t statistic, Student's t distribution, the one-sample t-test; [20min] the concept of two populations, two sample testing for mean differences, H_0 specification for left-sided, right-sided, two sided tests; [15min] the exact two-sample z-test under different variances and shared variance; [20min] the exact two-sample t-test under shared variance, the pooled standard deviation estimator

Lec 7 [25min] the approximate t-test (Welch-Satterthwaite approximation) under different variances, the Behrens-Fisher distribution; [5min] review of moments, definition of sample moments; [15min] the system of equations yielding the method of moments estimators (MME); [10min] MME for the expectation, MME for the variance; [20min] MME for the two parameters in the iid Binomial DGP, nonsensical MME values

Lec 8 [10min] MME for the iid Uniform DGP with one unknown endpoint parameter; [10min] definition of likelihood, the equivalence of the likelihood function with the JDF/JMF; [10min] definition of argmax, equivalence of argmax under strictly increasing functions; [10min] definition of maximum likelihood estimators (MLE) and estimates, the log likelihood; [10min] MLE for the iid Bernoulli DGP; [10min] MLEs for the iid Normal DGP; [10min] MLE for iid Uniform DGP with one unknown endpoint parameter; [5min] variances of MME and MLE for that cases

Lec 9 [5min] definition of relative efficiency and comparison of two estimators; [10min] the nonexistence of a minimum MSE estimator; [5min] definition of uniformly minimum variance unbiased estimators (UMVUEs); [10min] definition of the Cramer-Rao Lower Bound (CRLB), definition of Fisher Information; [55min] proof of the CRLB, definition of the score function, expectation of the score function is zero,

Lec 10 [10min] proof that the sample average is the UMVUE for the iid Bernoulli DGP; [15min] proof that the sample average is the UMVUE for the iid normal DGP; [10min] definition of asymptotically normal estimators; [15min] definition of consistent estimators, continuous mapping theorem, Slutsky's theorem; [10min] asymptotic normality of asymptotically normal estimators when using the estimator for its standard error; [5min] statement of main theorem for MMEs and MLEs: consistency, asymptotic normality, asymptotic efficience of the MLE

Lec 11 [5min] review of Taylor series; [40min] proof that the MLE is asymptotically normal and asmyptotically efficient; [5min] definition of the Wald test, the one-proportion z-test as a Wald test; [35min] the two-proportion z-test as a Wald test, the pooled proportion estimator

Lec 12 [5min] the one-sample t-test as an approximate one-sample z-test (Wald test); [10min] the approximate two-sample t-test as an approximate two-sample z-test (Wald test); [15min] derivation of the MLE, Fisher Information and a Wald test for the iid Gumbel DGP with known scale parameter; [15min] introduction to confidence sets, interval estimators, coverage probability; [20min] definition of the confidence interval (CI), CI construction via hypothesis test inversion

Lec 13 [10min] comparison of hypothesis testing with CI construction; [15min] approximate CIs for the iid normal DGP under the four assumptions; [10min] CI for one proportion; [15min] CI for the difference of two proportions; [10min] CIs for MLEs; [15min] proof that MSE improvements improve all three statistical inference goals, illustration of all three goals

Lec 14 [10min] meaninglessness of single inferences; [5min] odds-against reparameterization, odds-against point estimation; [15min] univariate delta method; [10min] CI for odds-against via delta method, CI for log-mean; [10min] risk ratio versus proportion difference; [20min] multivariate delta method; [5min] CI for the risk ratio

Lec 15 [10min] equivalence of the two-sided z and χ^2 tests, equivalence of the two-sided t and F tests; shape of the χ^2 and F distributions; [30min] χ^2 goodness of fit test for multinomial parameters, its H_0 and H_a, observed and expected counts, the test statistic; [40min] the χ^2 test of independence among two categorical variables, its H_0 and H_a, observed and expected counts, the test statistic

Lec 16 [35min] concept of assuming DGP models, discussion of what models are, discussion of many model candidates; [15min] model selection via largest likelihood, asymptotic bias of the log-likelihood estimator with substituted MLEs; [15min] the AIC metric, AIC model selection algorithm, penalizing complexity; [5min] the AICC metric

Lec 17 [5min] Akaike weights; [25min] statistical signifiance vs. clinical / practical significance of the effect; [5min] review of Type I/II errors; [40min] the multiple hypothesis testing / comparison problem, the 2×2 frequency table of test results, false discoveries,

Lec 18 [10min] definition of familywise error rate (FWER), weak FWER control; [10min] Bonferroni procedure; [15min] Sidak procedure; [25min] Simes procedure; [15min] definition of false discovery proportion, false discovery rate (FDR), setting for equivalence of FWER and FDR, statement that the Simes procedure provides strong control of the FDR

Lec 19 [10min] proof that p-values are uniformly distributed under H_0; [50min] derivation of the score test as a Wald test, the score test statistic for the logistic distribution with known scale parameter; [15min] definition of likelihood ratio (LR) test statistic, statement of its asymptotic convergence to a χ^2_1, LR test (LRT) derivation of the LR statistic for the iid Bernoulli DGP

Lec 20 [25min] Proof of the asymptotic convergence of the LR test statistic; [50min] generalized likelihood ratio test for reduced models nested in a full model, statement of asymptotic convergence to a χ^2 with degrees of freedom equal to the difference of number of parameters, demonstration that it differs from the goodness of fit test, demonstration in the normal iid DGP

Lec 21 [10min] visualizing the Wald, score and LR tests; [5min] testing entire DGPs; [10min] definition and illustration of the empirical CDF; [20min] one-sample Kolmogorov-Smirnov (KS) test, statement of Kolmogorov distribution and its critical values; [10min] tests of DGP equivalence among two populations; [15min] two-sample KS test

Lec 22 [45min] two-sample permutation test, partitioning the master population, dsicussion of possible test statistics, computational construction of its RET via resampling; [30min] the nonparametric bootstrap procedure, typical use cases, approximate CI construction, approximate hypothesis testing

Lec 23 [20min] causal inference vs statistical inference, counfounding variable, an illustration of one scenario with confounding and one scenario without confounding; [15min] definition of a two-arm treatment vs control experiment, definition of assignment / allocation / manipulation; [10min] the Rubin causal model, counterfactuals, additive treatment effect, selection bias inducing bias in the naive treatment estimator; [15min] response model linear in confounder and noise, noise as an approximate normal realization, the explicit bias term and its explanation; [20min] randomized experiments, bernoulli trial design, completely randomized trial design, statment that causal estimates are unbiased over errors and randomized assignments, statement that bias is small in any random assignment

This is more of a typical mathematics theory course than the rest of the data science series but it is still not like the other math courses you're used to; it will be more philosophical and very high in conceptual density while the mathematics itself is kept deliberately simple. 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.

The 650 Section

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

Course Materials

Textbook: I will be referencing Larry Wasserman's *All of Statistics: A concise course in statistical inference* which can be purchased on Amazon and Casella and Berger's *Statistical Inference* which can be purchased on Amazon. There is no excuse not to have these books. They are *required*. However, I will not ususally be teaching "from the book" — most of the material in the class comes from the lecture notes. The textbooks are a way to get "another take" on the material and they will only cover about only half of the material done in class. For the other half, you will have to make use of other resources. I also recommend Rice's *Mathematical Statistics and Data Analysis*, 3rd edition which can be purchased on Amazon as well but I will not reference it during class.

Computer Software: During lectures, there will be demoes 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. If `R` will be part of the homework, I will give you the 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.

Announcements

Course announcements will be made via slack in the #general channel. I am known to send a few slack messages per week on important issues. Thus, I will need the email address that you reliably check. The default email is whatever happens to be in CUNYfirst which many of you do not check. (See Homework #0 for more information).

Lectures on Zoom

Classes are 75 minutes and run from Wednesday, August 26 until Wednesday, December 9 for a total of 28 class meetings. However, only 23 of these will be lectures as two days are reserved for the two midterm exams (in class) and two meetings prior are in-class reviews. The exam schedule is given on page 10. **Zoom policies: your video must always be on. You can use an appropriate background but it must be a static image. No snap camera or the like. No chatting on zoom.**

Lecture Upload

As many previous students have noted, my handwritten notes are useful to me and not to many others. Thus, I will be rewarding students for taking notes, scanning them in and sending them to me. You will be rewarded in two ways: (1) if you do this for more than 10 lectures, you will be given the automatic 5 points (see grading policy on page 11) for your classroom participation grade and (2) you have the option for me to say your name publicly on the course homepage. Make sure you follow these instructions:

- You have *one week only* from the time of the lecture to email me lecture notes.
- There must be *one* file and it must be in PDF format *only*.
- The file must be named `lecxxkapelner.pdf` where you replace `xx` with two digits corresponding to the lecture number i.e. 01, 02, 09, 10, . . . , 23 and you replace `kapelner` with your last name in all lowercase letters. If your file is renamed incorrectly, I will tell you to rename it and send it back.
- The file must be <2MB. No exceptions. I will tell you to shrink the PDF and resend.
- Vertically oriented (readable without rotating your head 90 degrees.).
- You have to agree to the MIT license.

Homework

There will be 7–10 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. I will grade an *arbitrary subset of the assignment* which is determined after the homework is handed in. But you will still be penalized for leaving questions blank regardless of whichever subset I choose to grade.

During this pandemic, 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.

You are encouraged to seek help from me if you have questions. After class and during 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 2.5hr in-class time per week is 6-9 hours. I will aim for 6hr of homework per week on average. However, doing the homework well is your sole responsibility since by doing the homework you will study and understand the concepts in the lectures.

Late Homework

Late homework will be penalized 10 points per day 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 day of class, Wednesday, December 9. 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.

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

Homework LaTeX Bonus Points

Part of good mathematics is its beautiful presentation. Thus, **there will be a 1–7 point bonus** added to your homework grade for typing up your homework using the LaTeX typesetting 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 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.

Homework Extra Credit

There will be many extra credit questions sprinkled throughout the homeworks (although less for the 621 Masters students). 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 an A- based on these bonuses.

Homework #0

For your first homework (due immediately). You must:

(1) email me at kapelner@qc.cuny.edu from the email address you wish to be contacted at for this course (most commonly this is a gmail address) and in the email,

(2) you must say "My name is <Your Full Name as appears in the registrar>",

This constitutes a contract — you are agreeing to this syllabus.

I will email you back a password you can use to check the gradesly, the course grading site once the site is up (which should be a couple weeks into the semester).

This assignment is due Friday, August 28 at 5PM and will receive a grade of 0 or 100 with the usual 10 point penalty for lateness. If you took one of my classes before, I do not store your personal email address! You still have to do Homework #0.

Examinations

Examinations are solely based on homeworks! 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.

On zoom, the camera must be on your hands at all times. You may have to practice this before the exam.

Exam Schedule

See course homepage.

Exam 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. **No food, only drinks.** I do not recommend using pen but it if you must...

I also allow "cheat sheets" on examinations. For both midterms, you are allowed to bring one 8.5" $\times$ 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" $\times$ 11" sheet of paper (front and back). **Six sheets single sided are not allowed.**

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.

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. If you need to leave the country for an emergency, I will expect proper documentation as well.

Missing the Final

Automatic WU grade. You can get an F by coming and "taking" the final.

Special Services

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

Class Participation (and attendance)

I will be taking attendance (sometimes formally and sometimes informally) during the class. Attendance counts towards the class participation portion of your grade in equal part with how often you ask and answer questions during the lecture.

The Use of Slack as a Learning Management System

This class has a slack workspace (see page 1). As the course homepage is updated, 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). 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 so other students can see the Q&A. Slack is a wildly successful $17.2B company because businesses use it. Pretend you are working at one of these businesses: no posting about random stuff; keep things professional!

Slack will be setup about a week after class begins and you will get an email with instructions about how to sign up.

Grading and Grading Policy

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

Homework	20%
Class participation	5%
Midterm Examination I	20%
Midterm Examination II	20%
Final Examination	35%

The semester is split into three periods (1) from the beginning until midterm I (2) from midterm I to midterm II (3) from midterm II until the final. The material in each of the periods is tested evenly; thus, it counts 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 is curved and the curve will be quite generous. 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. To give an idea, of the students who finished the 368 course last time I taught it, there were 43% A's and 29% B's but I am under **no obligation** to repeat this curve.

Checking your grade and class standing

You can always check your grades in real-time using the grading site. You will enter in your QC ID number and the password I will provide to you after homework 0.

Auditing

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