Probability — The Science of Uncertainty and Data

MITx 6.431x

27/08/2018 – 23/12/2018

Sumário

1 Welcome to 6.431x (2018–08–27)

The course site is now open. We have released Unit 0, which introduces the course and summarizes the objectives and what you can expect to learn. It also contains lots of important information that you should read over carefully. We have also released an Entrance Survey, and will appreciate your help in improving this course. Unit 1 will be released next Monday.

This is a graduate level version of 6.041x, which has been offered several times, and we are once more excited to offer this material. We hope that you will find this course an enriching educational experience, helping you to master the fundamental concepts and tools of probability theory and its applications.

This is a *challenging* class. It is exactly at the same level, breadth, and depth as the corresponding residential MIT offering. MIT students typically *spend about 12 hours a week* on this subject, and you can expect to need a similar time commitment, perhaps even a bit more, depending on your background. But even if you do not have the time to do everything, you may still gain a lot by following just parts of the course.

We look forward to seeing you in class! And tell your friends about it!

Best wishes, Prof. John Tsitsiklis, Eren Kizildag (TA), and your course team

2 Unit 0: Overview

Unit 0 is the first unit available in the Courseware. It introduces the course and summarizes the objectives and what you can expect to learn. It also contains lots of important information that you should read over carefully.

Course overview; Course introduction, objectives, and study guide These sections introduce and overview the course and provide a guide for how to make the most of the wealth of materials that this course offers.

Syllabus, calendar, and grading policy Here you will find an outline of the units of this course, together with release and due dates. The same information is presented in a calendar format for your convenience. The grading policy is also explained in detail.

edX tutorial This sequence of videos gives a visual tutorial of how to use the basic elements of the edX platform.

Discussion forum and collaboration guidelines This section contains the course's guidelines for collaboration and using the dicussion forum. Please read them carefully and follow them throughout the course.

Homework mechanics and standard notation This section explains how to submit answers to problems and details the standard notation that should be used throughout the course when entering symbolic responses. Please read carefully and refer back to these documents when needed.

Textbook information This section describes how to access and navigate through the e-reader of excerpts from the course textbook. There is also information for purchasing a physical copy of the textbook as well as a link to textbook errata. While this textbook is recommended, the materials provided by this course are self-contained.

Micromasters, Certificates, and Honor Pledge This section provides information on how to earn a verified certificate for this course, as well as how to obtain the credential for the MITx Micromasters Program in Statistics and Data Science. You will also be asked to make a pledge to abide by the EdX Honor Code.

2.1 Course overview

2.1.1 Course character and objectives

Video: Course character and objectives

2.1.2 Why study probability?

Video: Why study probability?

2.1.3 Course contents

Video: Course contents

2.2 Course introduction, objectives and study guide

2.2.1 Introduction

Welcome to 6.431x, an introduction to probabilistic models, including random processes and the basic elements of statistical inference.

The world is full of uncertainty: accidents, storms, unruly financial markets, noisy communications. The world is also full of data. Probabilistic modeling and the related field of statistical inference are the keys to analyzing data and making scientifically sound predictions.

The course covers all of the basic probability concepts, including:

- multiple discrete or continuous random variables, expectations, and conditional distributions
- laws of large numbers

- the main tools of Bayesian inference methods
- an introduction to random processes (Poisson processes and Markov chains)

2.2.2 Course objectives

Upon successful completion of this course, you will:

At a conceptual level:

- Master the basic concepts associated with *probability models*.
- Be able to translate models described in words to mathematical ones.
- Understand the main concepts and assumptions underlying *Bayesian and classical inference*.
- Obtain some familiarity with the range of *applications of inference methods*.

At a more technical level:

- Become familiar with basic and common *probability distributions*.
- Learn how to use *conditioning* to simplify the analysis of complicated models.
- Have facility manipulating *probability mass functions*, *densities*, and *expectations*.
- Develop a solid understanding of the concept of *conditional expectation* and its role in inference.
- Understand the power of *laws of large numbers* and be able to use them when appropriate.
- Become familiar with the basic inference methodologies (for both *estimation* and *hypothesis testing*) and be able to apply them.
- Acquire a good understanding of two *basic stochastic processes* (*Bernoulli* and *Poisson*) and their use in modeling.
- Learn how to formulate simple dynamical models as *Markov chains* and analyze them.

2.2.3 Study guide

This class provides you with a great wealth of material, perhaps more than you can fully digest. This “guide"offers some tips about how to use this material.

Start with the overview of a unit, when available. This will help you get an overview of what is to happen next. Similarly, at the end of a unit, watch the unit summary to consolidate your understanding of the “big picture"and of the relation between different concepts.

Watch the lecture videos. You may want to download the slides (clean or annotated) at the beginning of each lecture, especially if you cannot receive high-quality streaming video. Some of the lecture clips proceed at a moderate speed. Whenever you feel comfortable, you may want to speed up the video and run it faster, at 1.5x.

Do the exercises! The exercises that follow most of the lecture clips are a most critical part of this class. Some of the exercises are simple adaptations of you may have just heard. Other exercises will require more thought. Do your best to solve them right after each clip — do not defer this for later — so that you can consolidate your understanding. After your attempt, whether successful or not, do look at the solutions, which you will be able to see as soon as you submit your own answers.

Solved problems and additional materials. In most of the units, we are providing you with many problems that are solved by members of our staff. We provide both video clips and written solutions. Depending on your learning style, you may pick and choose which format to focus on. But in either case, it is important that you get exposed to a large number of problems.

The textbook. If you have access to the textbook, you can find more precise statements of what was discussed in lecture, additional facts, as well as several examples. While the textbook is recommended, the materials provided by this course are self-contained. See the "Textbook information"tab in Unit 0 for more details.

Problem sets. One can really master the subject only by solving problems — a large number of them. Some of the problems will be straightforward applications of what you have learned. A few of them will be more challenging. Do not despair if you cannot solve a problem – no one is expected to do everything perfectly. However, once the problem set solutions are released (which will happen on the due date of the problem set), make sure to go over the solutions to those problems that you could not solve correctly.

Exams. The midterm exams are designed so that in an on-campus version, learners would be given two hours. The final exam is designed so that in an on-campus version, learners would be given three hours. You should not expect to spend much more than this amount of time on them. In this respect, those weeks that have exams (and no problem sets!) will not have higher demands on your time. The level of difficulty of exam questions will be somewhere between the lecture exercises and homework problems.

Time management. The corresponding on-campus class is designed so that students with appropriate prerequisites spend about 12 hours each week on

lectures, recitations, readings, and homework. You should expect a comparable effort, or more if you need to catch up on background material. In a typical week, there will be 2 hours of lecture clips, but it might take you 4–5 hours when you add the time spent on exercises. Plan to spend another 3–4 hours watching solved problems and additional materials, and on textbook readings. Finally, expect about 4 hours spent on the weekly problem sets.

Additional practice problems. For those of you who wish to dive even deeper into the subject, you can find a good collection of problems at the end of each chapter of the print edition of the book, whose solutions are available online.

2.3 Syllabus, calendar, and grading policy

2.3.1 Syllabus

6.431x Fall 2018 Syllabus

- Unit 0: Overview (released Tue. August 28)
- Unit 1: Probability models and axioms (released Mon. Sep 3; Sections 1.1–1.2)
  - L1: Probability models and axioms
  - Problem Set 1 due on Tue Sept 11
- Unit 2: Conditioning and independence (released Mon. Sept 10; Sections 1.3–1.5)
  - L2: Conditioning and Bayes' rule
  - L3: Independence
  - Problem Set 2 due on Tue Sept 18
- Unit 3: Counting (released Mon. Sept 17; Section 1.6)
  - L4: Counting
  - Problem Set 3 due on Tue Sept 25
- Unit 4: Discrete random variables (released Wed. Sept 19; Sections 2.1–2.7)
  - L5: Probability mass functions and expectations
  - L6: Variance; Conditioning on an event; Multiple r.v.'s
  - L7: Conditioning on a random variable; Independence of r.v.'s
  - Problem Set 4 due on Tue Oct 2
- Exam 1 (Timed) : Covers material from L1 to L7 (released Wed. Oct 3; due on Tue. Oct 9)
- Unit 5: Continuous random variables (released Mon. Oct 1; Sections 3.1–3.5)
  - L8: Probability density functions
  - L9: Conditioning on an event; Multiple r.v.'s

  - L10: Conditioning on a random variable; Independence; Bayes' rule
  - Problem Set 5 due on Tue. Oct 16

- Unit 6: Further topics on random variables (released Mon. Oct 15; Sections 4.1–4.3, 4.5)
  - L11: Derived distributions
  - L12: Sums of r.v.'s; Covariance and correlation
  - L13: Conditional expectation and variance revisited; Sum of a random number of r.v.'s
  - Problem Set 6 due on Tue. Oct 23

- Unit 7: Bayesian inference (released Mon. Oct 22 Sections 3.6, 8.1–8.4)
  - L14: Introduction to Bayesian inference
  - L15: Linear models with normal noise
  - L16: Least mean squares (LMS) estimation
  - L17: Linear least mean squares (LLMS) estimation
  - Problem Set 7a due on Tue. Oct 30
  - Problem Set 7b due on Tue. Nov 6

- Exam 2 (Timed): Covers material from L8 to L17 (released Wed. Nov 1; due on Nov 13)
- Unit 8: Limit theorems and classical statistics (released Mon. Nov 5; Sections 5.1–5.4, pp. 466–475)
  - L18: Inequalities, convergence, and the Weak Law of Large Numbers
  - L19: The Central Limit Theorem (CLT)
  - L20: An introduction to classical statistics
  - Problem Set 8 due on Tue. Nov 27

- Unit 9: Bernoulli and Poisson processes (released Tue. Nov 14; Sections 6.1-6-2)
  - L21: The Bernoulli process
  - L22: The Poisson process
  - L23: More on the Poisson process
  - Problem Set 9 due on Tue. Dec 4

- Unit 10: Markov chains (released Tue. Nov 26; Sections 7.1–7.4)
  - L24: Finite-state Markov chains
  - L25: Steady-state behavior of Markov chains
  - L26: Absorption probabilities and expected time to absorption
  - Problem Set 10 due on Tue. Dec 11

- Final Exam (Timed) (released Wed. Dec 12; due on Sun. Dec 23)

Note: Problem set and exam due dates are at the end of the specified date, at 23:59 UTC.

2.3.2 Calendar

6.431x Fall 2018 Calendar

MONDAY	TUESDAY	WEDNESDAY
9/3 Unit 1 released: Probability models and axioms (Secs. 1.1-1.2)	**9/4**	**9/5**
9/10 Unit 2 released: Conditioning and independence (Secs. 1.3-1.5)	**9/11** Problem Set 1 due	**9/12**
9/17 Unit 3 released: Counting (Sec. 1.6)	**9/18** Problem Set 2 due	**9/19** Unit 4 released: Discrete r.v.'s (Ch. 2)
9/24	**9/25** Problem Set 3 due	**9/26**
10/1 Unit 5 released: Continuous r.v.'s (Secs. 3.1-3.5)	**10/2** Problem Set 4 due	**10/3** Exam 1 (Timed) released
10/8	**10/9** Exam 1 (Timed) due	**10/10**
10/15 Unit 6 released: Further topics on r.v.'s (Secs. 4.1-4.3, 4.5)	**10/16** Problem Set 5 due	**10/17**
10/22 Unit 7 released: Bayesian inference (Secs. 3.6, 8.1-8.4)	**10/23** Problem Set 6 due	**10/24**
10/29	**10/30** Problem Set 7a due	**10/31**
11/5 Unit 8 released: Limit theorems and classical statistics (Secs. 5.1-5.4, pp. 466-475)	**11/6** Problem Set 7b due	**11/7** Exam 2 (Timed) released
11/12	**11/13** Exam 2 (Timed) due	**11/14** Unit 9 released: Bernoulli and Poisson processes (Secs. 6.1-6.-2)
11/19	**11/20**	**11/21**
11/26 Unit 10 released: Markov chains (Secs. 7.1-7.4)	**11/27** Problem Set 8 due	**11/28**
12/3	**12/4** Problem Set 9 due	**12/5**
12/10	**12/11** Problem Set 10 due Final Exam (Timed) released	**12/12**
12/17	**12/18**	**12/19** 12/20 Final Exam (Timed) due

Notes:

- The due dates for the weekly problem sets and the exams are fixed and cannot be changed or modified for any individuals. Please plan accordingly.
- Problem set and exam due dates are at the end of the specified date, at 23:59 UTC.
- The calendar above shows only Tuesdays, Wednesdays, and Thursdays, since these are the only days of the week when materials will be released or due, except the final exam.

2.3.3 Gradingp policy

Grading policy Your overall score in this class will be a weighted average of your scores for the different components, with the following weights:

- 20% for the lecture exercises (divided equally among the 26 lectures)
- 40% for the problem sets (divided equally among 11 problem sets)
- 12% for the first midterm exam (timed)
- 12% for the second midterm exam (timed)
- 16% for the final exam (timed)

To earn a verified certificate for this course, you will need to obtain an *overall score* of 60% or more of the maximum possible overall score.

Note that not every problem set or set of lecture exercises will have the same number of raw points. For example, Problem Set 1 may have 30 points and Problem Set 2 may have 35 points. However, each one receives the same weight for the purposes of calculating your overall score.

As an illustrative example, if you receive 20 points out of 30 on Problem Set 1, this will contribute [Math Processing Error] to your overall score. Similarly, if you receive 30 points out of 35 on Problem Set 2, this will contribute [Math Processing Error] to your overall score.

Under the “Progress"tab at the top, you can see your score broken down for each assignment, as well as a summary plot.

Timed Exams The 2 midterm exams and one final exam are *timed exams*. This means that each exam is available for approximately a week, but once you open the exam, there is a limited amount of time (48 hours), counting from when you start, within which you must complete the exam. Please plan in advance for the exams. If you do not complete the whole exam during the allowed time, you will miss the points associated with the questions that have not been answered. The exams are designed to assess your knowledge. There are no extensions granted to these deadlines. You can find the exam dates on the calendar on the previous page. Note that the timed exams cannot be completed using the edX mobile app.

MITx Committment to Accessibility If you have a disability-related request regarding accessing an MITx course, including exams, please contact the course team as early in the course as possible (at least 2 weeks in advance of exams opening) to allow us time to respond in advance of course deadlines. Requests are reviewed via an interactive process to meet accessibility requirements for learners with disabilities and uphold the academic integrity for MITx.

2.4 edX Tutorial

2.4.1 Basics

Video: edX Basics

2.4.2 Courseware navigation

Video: Courseware navigation

2.4.3 Top-level navigation

Video: Top-level navigation

2.4.4 Discussion forums

Video: Discussion forums

2.4.5 Summary

Video: Summary

2.5 Discussion forum and collaboration guidelines

2.6 Discussion forum guidelines

Discussion forum overview The course provides an online discussion forum for you to communicate with the course team and other learners. You may access the forum through the “Discussion"tab at the top of the page, as well as through many embedded discussions within each unit. We recommend using the embedded discussions within each unit to discuss topics related to a specific unit’s materials, whether it’s lectures, solved problems, or problem set problems. Please see the guidelines below for more information on how to use these embedded discussions.

For other more general discussions, you may use the “Discussion"tab at the top of the page. When creating a new post, *please choose one of the following categories that best describes your post*:

- *Introductions*: Introduce yourself to your fellow learners and find out more about them!

- *Micromasters*: Ask questions related to the MITx Micromaster Program in Statistics and Data Science and meet other Micromasters fellows!
- *Course Feedback*: Let the course team know how you are finding the course, what you think works well, and what you would like to see improved.
- *Technical Problems*: Let the course team know about any technical issues you are dealing with (e.g., playing videos, entering answers, etc).
- *General*: Other general discussions.

Discussion forum guidelines The discussion forum is the main way for you to communicate with the course team and other learners. We hope it contributes to a sense of community and serves as a useful resource for your learning. Here are some guidelines to help you successfully navigate and interact on the forum:

- *Use discussion while working through the material*. Beginning with Unit 1, each lecture will contain an embedded discussion located at the bottom of the lecture overview clip, which is the first or second clip of that lecture sequence. You should discuss anything related to that lecture's video clips or exercises there. Click "Show Discussion"to see all discussions associated with the lecture, and click "Add a Post"to post a new topic. In addition, every solved problem and problem set problem will have its own embedded discussion located at the bottom of their respective pages. As with the lecture discussions, click "Show Discussion"and "Add a Post"to see and create discussion topics related to that specific problem. We recommend that you use these in-page discussion boards to help focus discussions on specific topics.
- *Use informative topic titles and tags*. To make it easier to identify relevant discussion topics, please use informative titles and tags when creating a new discussion topic. We suggest using titles or tags that are as informative as possible, e.g., "lecture X, exercise Y on topic W, clarify part Z"
- *Be very specific*. Provide as much information as possible about what you need help for: Which part of what problem or video? Why do you not understand the question? Do you need help understanding a particular concept? What have you tried doing so far? Use a descriptive title to your post. This will attract the attention of other learners having the same issue.
- *Observe the honor code*. We encourage collaboration and help, but please do not ask for nor post problem solutions.
- *Upvote good posts*. This applies to questions and answers. Click on the green plus button so that good posts can be found more easily.
- *Search before asking*. The forum can become hard to use if there are too many threads, and good discussions happen when people participate in the same thread. Before asking a question, use the search feature by clicking on the magnifying glass on the left-hand side.

- *Write clearly*. We know that English is a second language for many of you but correct grammar will help others to respond. Avoid ALL CAPS, abbrv of wrds (abbreviating words), and excessive punctuation!!!!

Please Introduce Yourself! Let's get started by introducing yourselves on the discussion forum. A lot of the learning in this class will happen in your interactions with each other. Click on the post titled "Introduce yourself!"below, and respond to it by telling everyone your name, where you are from, why you are taking this course, and whatever else you would like to share! Your post will be indexed in the "Introductions"category in the forum.

2.7 Collaboration guidelines

We encourage you to interact with your fellow learners and engage in active discussion about the course. Please use the guidelines below for acceptable collaboration. The staff will be proactive in removing posts and replies in the discussion forum that have stepped over the line.

- Given a problem, it is ok to discuss the general approach to solving the problem.
- You can work jointly to come up with the general steps for the solution.
- It is ok to get a hint, or several hints for that matter, if you get stuck while solving a problem.
- You should work out the details of the solution yourself.
- It is not ok to take someone else's solution and simply copy the answers from their solution into your checkboxes.
- It is not ok to take someone else's formula and plug in your own numbers to get the final answer.
- It is not ok to post answers to homework and lab problems before the submission deadline.
- It is not ok to look at a full step-by-step solution to a problem before the submission deadline.
- It is ok to have someone show you a few steps of a solution where you have been stuck for a while, provided of course, you have attempted to solve it yourself without success.
- After you have collaborated with others in generating a correct solution, a good test to see if you were engaged in acceptable collaboration is to make sure that you are able to do the problem on your own.

2.8 Homework mechanics and standard notation

2.8.1 Checking and submitting an answer

Checking and submitting an answer For each problem, you will have between 2 to 5 attempts to submit an answer, with the exception of problems where an

attempt essentially reveals the answer (e.g., True/False questions), for which you will be limited to a single attempt.

To submit your answer, click the “Submit "button. This will automatically submit the problem for grading purposes, and the edX platform is able to verify your answer and give you immediate feedback as to whether or not your answer is correct. To save your answer without submitting it for grading purposes, click the “Save"button. Your answer will be restored when you return to the problem.

The number of attempts allowed as well as the number of attempts you’ve already made will always be visible on a problem’s page at the bottom, next to the “Check"button. Please note that for problems consisting of multiple parts, hitting the button will count as an attempt for all parts of the problem. Unfortunately, it is not possible to submit answers for one part at a time.

For lecture exercises, a “Show Answer(s)"button will appear immediately after you submit the correct answer or use all of your attempts. Clicking this button will reveal the correct answers and solutions.

For homework problems, the “Show Answer(s)"button will appear after the due date of the homework.

You are strongly encouraged to look at the solutions even if your answer is correct.

Answer formats This course will use several answer formats:

- Multiple choice: Select the correct option from the dropdown menu or radio buttons.
- Numerical answers: Enter a number, either in decimal (e.g., ’3.14159’) or fractional form (e.g., ’22/7’). Do not enter any non-numerical letters or symbols. To account for rounding, the system will accept a range of answers as correct. Unless otherwise specified in the problem, the default tolerance range will be +/-3% of the correct answer.
- Symbolic answers: Some problems will ask for a symbolic answer (e.g., ’n*(n+1)/2’). See the next section on “Standard notation"for details on how to submit such answers.

Below are some example problems for you to familiarize yourselves with how these problem types work with different number of attempts. These problems are not graded and have no impact on your grade.

2.9 Textbook information

2.10 Entrance survey