Probability:
The Science of Uncertainty and Data

MITx 6.431x

2018/08/28 – 2018/12/23

Sumário

1 Welcome to 6.431x (2018/08/28)

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 (2018/08/28)

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 (transcripts, slides)

2.1.2 Why study probability?

Video: Why study probability? (transcripts, slides)

2.1.3 Course contents

Video: Course contents (transcripts, slides)

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)
- 20% for the problem sets (divided equally among 11 problem sets)
- 18% for the first midterm exam (timed)
- 18% for the second midterm exam (timed)
- 24% 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 $\frac{20}{30} \times \frac{20\%}{11} = 1.21\%$ to your overall score. Similarly, if you receive 30 points out of 35 on Problem Set 2, this will contribute $\frac{30}{35} \times \frac{20\%}{11} = 1.56\%$ 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

2.5.1 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.

2.5.2 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.

2.6 Homework mechanics and standard notation

2.6.1 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.

2.6.2 Standard notation

Many exercises and problems throughout the course will ask you to provide an algebraic answer in terms of symbols. Please follow the guidelines below when entering your responses. Below your answer textbox, the system will also display, in a "pretty"format, what it has interpreted your input to be. However, this display is not perfect (for example, it does not catch all cases of missing close parentheses) so please also check your text input carefully.

- Symbols are case-sensitive: a and A are different — make sure to use the correct case as specified in the problem
- Parentheses: make sure that your parentheses are properly balanced — each open parenthesis should have a matching close parenthesis!
- Elementary arithmetic operations: use the symbols + , - , * , / for addition, subtraction, multiplication, and division, respectively
- $1 + bc - d/e$ should be entered as 1+b*c-d/e
- For multiplication, use * explicitly:
  - in the example above, enter b*c; do NOT enter bc
  - for $2n(n + 1)$, enter 2*n*(n+1); do NOT enter 2n(n+1)
  - although the "pretty"display underneath your answer looks correct if you do not include * s, your answer will be marked incorrect!
- Exponents: use the symbol ^ to denote exponentiation
  - 2^n should be entered as 2^n
  - x^{n+1} should be entered as x^(n+1)
- Square root: use the string of letters sqrt , followed by enclosing what is under the square root in parentheses
  - $\sqrt{-1}$ should be entered as sqrt(-1)
- Mathematical constants: use the symbol e for the base of the natural logarithm, e; use the string of letters pi for π
  - $e^{i\pi} + 1$ should be entered as e^(i*(pi))+1
- Order of operations: 1) parentheses, 2) exponents and roots, 3) multiplication and division, 4) addition and subtraction
  - $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ should be entered as (1/sqrt(2*(pi)))*e^(-(x^2)/2)
  - a/b*c is interpreted as $\frac{a}{b}c$
  - enter a/(b*c) for $\frac{a}{bc}$
  - When in doubt, use additional parentheses to remove possible ambiguitites
- Natural logarithm: although in lectures and solved problems we will sometimes use the notation "log"(instead of "ln"), you should use the string of letters ln , followed by the argument enclosed in parentheses
  - $\ln(2x)$ should be entered as ln(2*x)
- Trigonometric functions: use the usual 3-letter symbols to denote the standard trigonometric functions
  - $\sin(x)$ should be entered as sin(x)
- Greek letters: use the Latin-character name to denote each Greek letter
  - $\lambda e^{-\lambda t}$ should be entered as lambda*e^(-lambda*t)
  - $\mu\alpha\theta$ should be entered as mu*alpha*theta

- Factorials, permutations, combinations: you will not need enter these for any symbolic answers; do NOT use ! in your answers as it will not be evaluated correctly!

Figura 1: Standard Notation Summary: 1

Symbols	These are case sensitive. Use the correct case as specified in the problem.	**n** and **N** are different.	Do NOT enter x for X
Parentheses	Match each open parenthesis with a close parenthesis.		
Elementary Arithmetic Operations	Use the symbols **+,-,*,/** for addition, subtraction, multiplication, and division, respectively.	Enter **1+b*c-d/e** for $1+bc-d/e$	
	For multiplication, use * explicitly. Although the "pretty" display underneath your answer looks correct if you do not include * s, your answer will be marked incorrect!	Enter **b*c** for bc in the example above Enter **2*n*(n+1)** for $2n(n+1)$	Do NOT enter bc for bc Do NOT enter 2n(n+1) for $2n(n+1)$
Exponents	Use the symbol **^** to denote exponentiation.	Enter **2^n** for 2^n Enter **x^(n+1)** for x^{n+1}	
Square Root	use the string of letters **sqrt**, followed by enclosing what is under the square root in parentheses.	Enter **sqrt(-1)** for $\sqrt{-1}$	

Figura 2: Standard Notation Summary: 2

Mathematical Constants	Use the symbol **e** for the base of the natural logarithm, e. Use the string of letters **pi** for π.	Enter **e^(i*(pi))+1** for $e^{i\pi}+1$	
Order Of Operations	1) parentheses 2) exponents and roots 3) multiplication and division 4) addition and subtraction When in doubt, use additional parentheses to remove possible ambiguitites.	Enter **a/b*c** for $\frac{a}{b}\cdot c$ Enter **a/(b*c)** for $\frac{a}{bc}$ Enter **(1/sqrt(2*(pi)))*e^(-(x^2)/2)** for $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	Do NOT enter a/b*c for $\frac{a}{bc}$
Natural Logarithm	Although in lectures and solved problems we will sometimes use the notation "log" (instead of "ln"), you should use the string of letters **ln**, followed by the argument enclosed in parentheses.	Enter **ln(2*x)** for $\ln(2x)$	Do NOT enter log(2*x) for $\ln(2x)$
Trigonometric Functions	Use the usual 3-letter symbols to denote the standard trigonometric functions	Enter **sin(x)** for $\sin(x)$	Do NOT enter sin x for $\sin(x)$
Greek Letters	Use the Latin-character name to denote each Greek letter	Enter **lambda*e^(-lambda*t)** for $\lambda e^{-\lambda t}$ Enter **mu*alpha*theta** for $mu\alpha\theta$	
Factorials, Permutations, Combinations	You will not need enter these for any symbolic answers.		Do NOT use ! in your answers as it will not be evaluated correctly!

2.7 Textbook information

2.7.1 Textbook

The class follows closely the text Introduction to Probability, 2nd edition, by Bertsekas and Tsitsiklis, Athena Scientific, 2008; see the publisher's website or Amazon.com for more information.

While this textbook is recommended, the materials provided by this course are self-contained. Furthermore, the publisher has made available, for the purposes of this class, the summary tables that are included in the text. These can be found under the "Resources"tab, or directly by following this link. In various places within the courseware, there will also be links to specific sections and pages to the excerpts from the textbook relevant to the material at hand. These links will also take you to the e-reader, jumping directly to the specific sections and pages.

To adjust the zoom level in the e-reader, click the '+' and '-' buttons at the top-right to zoom in and out, respectively. Or, choose a specific zoom level using the drop-down menu. Depending on your operating system and web browser, you may encounter occasional artifacts or imperfect rendering of some formulas. Please try adjusting the zoom level to find the one that gives the best readability. We recommend using Firefox as it renders the text most accurately.

2.7.2 Ordering and other information

The class follows closely the text Introduction to Probability, 2nd Ed., by Bertsekas and Tsitsiklis. If interested in purchasing a copy of the textbook, it is available through Amazon.

Textbook errata can be found here. Ignore "Corrections to the 1st and 2nd printing."These do not apply to the currently available printed version.

2.8 Micromasters, Certification, and Honor Pledge

2.8.1 Micromasters

Video: MITs Micromasters in Statistics and Data Science

This course is part of the MITx Micromaster Program in Statistics and Data Science. Welcome to the program!

About the Program The MITx Micromasters program in Statistics and Data Science is comprised of four EdX courses and a virtually proctored exam that will provide you with the foundational knowledge essential to understanding the methods and tools used in data science, and hands-on training in data analysis and machine learning. You will dive into the fundamentals of probability and statistics, as well as learn, implement, and experiment with data analysis techniques and machine learning algorithms. This program will prepare you to become an informed and effective practitioner of data science who adds value to an organization and will also accelerate your path towards an MIT PhD or a Master's at other universities.

Anyone can enroll in the Micromasters program just as in any EdX courses. It is designed for learners who want to acquire sophisticated and rigorous training in data science without leaving their day job but also without compromising quality. There is no application process. To excel in the entire program, make sure you learn the foundational material covered in this course. You will also need some knowledge of matrices and proficiency in Python programming.

What You'll Learn You'll lean about:

- Master the foundations of data science, statistics, and machine learning
- Analyze big data and make data-driven predictions through probabilistic modeling and statistical inference; identify and deploy appropriate modeling and methodologies in order to extract meaningful information for decision making
- Develop and build machine learning algorithms to extract meaningful information from seemingly unstructured data; learn popular unsupervised learning methods, including clustering methodologies and supervised methods such as deep neural networks
- Master techniques in modern data analysis to leverage big datasets; use python and R skillfully to analyze data

How to earn the Micromasters credential To earn the MITx Micromasters credential in statistics and data science, you must successfully pass and receive a Verified Certificate in each of the 4 courses listed below and pass the final Capstone Exam:

- 6.431x Probability–the Science of Uncertainty and Data
- 14.310Fx Data Analysis in Social Sciences
- 18.6501x Fundamentals of Statistics
- 6.86x Machine Learning with Python–From Linear Models to Deep Learning
- DS-CFx Captone Exam in Statistics and Data Science

All the courses are taught by MIT faculty at a similar pace and level of rigor as an on-campus course at MIT.

More information If you are interested in the Micromasters program, visit https://www.edx.org/micromasters/mitx-statistics-and-data-science. For more detail on this program and credit pathways, please visit the MITx Micromasters Portal, which includes a "Contact us"link at the very bottom left. You may also find the FAQ helpful. Finally, you can start connecting with fellow Micromasters learners on the discussion forum!

2.8.2 Certification

To earn a Verified Certificate in this course, you need to:

- Upgrade your status to be a verified learner - The fee is $300.
- Verify your identity - ID Verification.
- Pass the course - at least 60% on your final grade.

You have limited time to switch to a Verified Certificate learner – you should get ID Verified as soon as you register as a Verified learner. See the EdX FAQ for more details on certificates.

Note: It is your responsibility to make sure that your ID verification is valid during the whole course.

A verified certificate indicates that you have successfully completed the course, but will not include a specific grade. Certificates are issued by edX under the name of MITx and are delivered online through your dashboard on edx.org.

2.8.3 EdX Honor Code Pledge

By enrolling in an EdX course, you have already agreed with the EdX Honor Code, which means that you will do the following:

- Complete all graded material (graded assignments and exams) with your own work and only your own work. You will not submit the work of any other person or have anyone else submit work under your name.
- Maintain only one user account and not let anyone else use your username and/or password. Having two user accounts registered in this course will constitute cheating. Not engage in any activity that would dishonestly improve your results, or improve or hurt the results of others."."Not collaborate with anyone other than staff on the exam questions. This means comparing answers, working as teams, or sharing answers in any way.
- Not post answers to any problems that are used to assess learner performance.
- Always be polite and respectful when communicating across the platform (with other learners and the staff).

We will strictly enforce this honor code pledge. Learners found violating this pledge will be dealt with directly. If we become aware of any suspicious activity we reserve the right to remove credit, not award a certificate, revoke a certificate, ban from this and other courses in the MITs Micromasters Program in Statistics and Data Science as well as notify edX for other actions. We take academic honesty very, very seriously at MIT. With the introduction of the Micromasters Credential, the importance of honesty in work has been elevated to a much higher level than before. We will diligently monitor this and be very proactive.

2.9 Entrance survey

For us to offer the best course experience possible, we'd like to ask you to answer a few questions about yourself: Entrance Survey.

3 Unit 1: Probability models and axioms (2018/09/03)

3.1 Lecture 1: probability models and axioms

Atention: Exercises due Sep 11, 2018 20:59:59 -03.

3.1.1 Motivation

Video: Motivation

Let's face it. Life is uncertain. But one thing is certain. We need a way to make predictions and make decisions under uncertainty. Probabilistic models can help you answer questions, such as:

- What are the odds that there will be a long line at the supermarket checkout counter?
- How likely is it that my GPS device is off by more than 10 meters?
- What are the odds that I will have a car accident next year?
- How likely is it that the air traffic control radar will miss the approaching plane?
- Should I invest in the stock market now or wait?
- Can I use a probabilistic model of social networking data to create a marketing campaign?
- How do we use a statistical model to decide if a medical treatment is effective?
- How do I model the huge amounts of data that are now becoming available in so many different fields, big data, as they call it, and extract useful information?

I am *John Tsitsiklis*. And I'm *Patrick Jaillet*. Our mission in this class is to give you the tools to model and analyze uncertain situations no matter what your discipline.

To do that, we will use the language and precision of mathematics, but we will also build your intuition. This is an ambitious class. The online version is at *the same level as the one offered to MIT students*.

It covers a lot of material. Beyond the basics, you will learn about random processes and about extracting information from data. In the end, you will be able to make much better sense of the uncertainty around you. The rewards are certain to come.

Let's face it. Life is uncertain.

3.1.2 Overview and slides

This lecture sequence introduces the basic structure of probability models, including the sample space and the axioms that any probabilistic model should obey, together with some consequences of the axioms and some simple examples.

Video: Lecture 01: Probability Models and Acioms (transcripts, slides, annotated slides)[1] [2] [3]

Welcome to the first lecture of this class. You may be used to having a first lecture devoted to general comments and motivating examples. This one will be different.

We will dive into the heart of the subject right away. In fact, today we will accomplish a lot. By the end of this lecture, you will *know about all of the elements of a probabilistic model*.

A probabilistic model is a quantitative description of a situation, a phenomenon, or an experiment whose outcome is uncertain. Putting together such a model involves two key steps:

- First, we need to describe the possible outcomes of the experiment. This is done by specifying a so-called *sample space*.

- Second, we specify a *probability law*, which assigns probabilities to outcomes or to collections of outcomes.

The probability law tells us, for example, whether one outcome is much more likely than some other outcome.

Probabilities have to satisfy certain basic properties in order to be meaningful. These are the *axioms of probability theory*. For example probabilities cannot be negative. Interestingly, there will be very few axioms, but they are powerful, and we will see that they have lots of consequences. We will see that they imply many other properties that were not part of the axioms.

We will then go through a couple of very simple examples involving models with either *discrete* or *continuous* outcomes. As you will be seeing many times in this class, discrete models are conceptually much easier. Continuous models involve some more sophisticated concepts, and we will point out some of the subtle issues that arise. And finally, we will talk a little bit about the big picture, about the role of probability theory, and its relation with the real world.

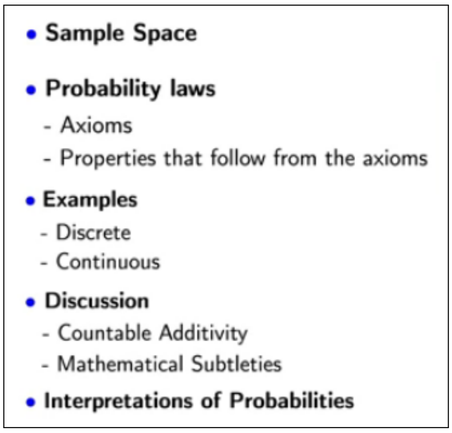
[1] The same material, in live lecture hall format, can be found here and here.

[2] You can also take this occasion to review some concepts related to sets (especially De Morgan's laws), sequences, and infinite series, by watching the "Mathematical background"sequence of clips.

[3] More information is given in the text:

- Sets: Section 1.1
- Probabilistic models: Section 1.2

Figura 3: Objectives of Lecture 1



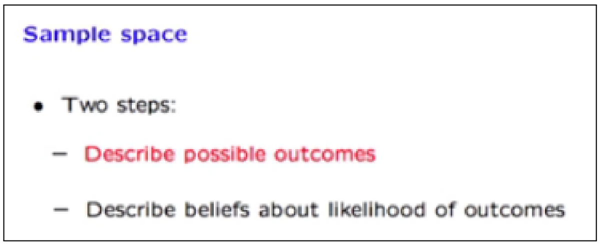
3.1.3 Sample space

Video: Sample space (transcript)

Putting together a probabilistic model — that is, a model of a random phenomenon or a random experiment — involves two steps.

- First step, we describe the *possible outcomes* of the phenomenon or experiment of interest.
- Second step, we describe our beliefs about the *likelihood of the different possible outcomes* by specifying a *probability law*.

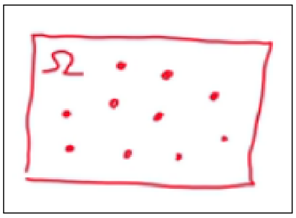
Here, we start by just talking about the first step, namely, the description of the possible outcomes of the experiment.



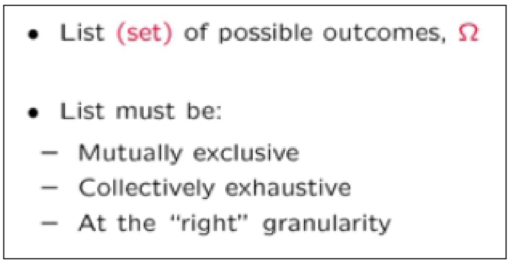
So we carry out an experiment. For example, we flip a coin. Or maybe we flip five coins simultaneously. Or maybe we roll a die.

Whatever that experiment is, it has a number of possible outcomes, and we start by *making a list of the possible outcomes* — or, a better word, instead of the word "list", is to use the word "set", which has a more formal mathematical meaning.

So we create a set that we usually denote by capital omega, Ω. That set is called the *sample space* and is the set of **all possible outcomes of our experiment**:



The elements of that set should have certain properties. Namely, the elements should be **mutually exclusive** and **collectively exhaustive**.



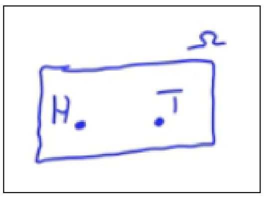
What does that mean? Mutually exclusive means that, if at the end of the experiment, I tell you that this outcome happened, then it should not be possible that this outcome also happened. At the end of the experiment, there can *only be one of the outcomes that has happened*.

Being collectively exhaustive means something else — that, together, all of these *elements of the set exhaust all the possibilities*. So no matter what, at the end, you will be able to point to one of the outcomes and say, that's the one that occurred.

To summarize: this set should be such that, at the end of the experiment, you should be always able to *point to one, and exactly one, of the possible outcomes* and say that this is the outcome that occurred.

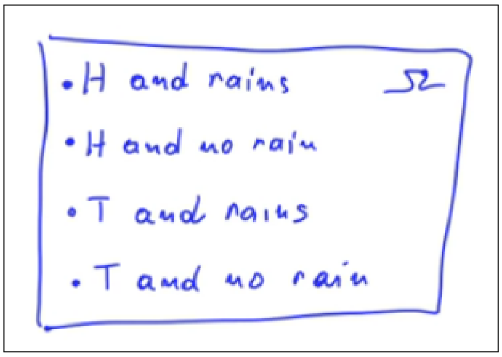
Physically *different outcomes should be distinguished in the sample space* and correspond to distinct points. But when we say physically different outcomes, what do we mean? We really mean *different in all relevant aspects* but perhaps not different in irrelevant aspects. Let's make more precise what I mean by that by looking at a very simple, and maybe silly, example, which is the following.

Suppose that you flip a coin and you see whether it resulted in heads or tails. So you have a perfectly legitimate sample space for this experiment which consists of just two points — heads and tails:



Together these two outcomes *exhaust all possibilities*. And the two outcomes are *mutually exclusive*. So this is a very legitimate sample space for this experiment.

Now suppose that while you were flipping the coin, you also looked outside the window to check the weather. And then you could say that my sample space is really, heads, and it's raining. Another possible outcome is heads and no rain. Another possible outcome is tails, and it's raining, and, finally, another possible outcome is tails and no rain.



This set, consisting of four elements, is also a perfectly legitimate sample space for the experiment of flipping a coin. The elements of this sample space are mutually exclusive and collectively exhaustive. Exactly one of these outcomes is going to be true, or will have materialized, at the end of the experiment.

So which sample space is the correct one? This sample space, the second one, involves some *irrelevant details*. So the preferred sample space for describing the flipping of a coin, the preferred sample space is the simpler one, the first one, which is sort of at the *right granularity*, **given what we're interested in**.

But ultimately, the question of which one is *the right sample space depends on what kind of questions you want to answer*. For example, if you have a theory that the weather affects the behavior of coins, then, in order to play with that theory, or maybe check it out, and so on, then, in such a case, you might want to work with the second sample space.

This is a common feature in all of science. Whenever you put together a model, you need to decide how detailed you want your model to be. And *the right level of detail is the one that captures those aspects that are relevant and of interest to you*.

3.1.4 Exercise: Sample Space

Exercise 3.1.4-1: Sample space

For the experiment of flipping a coin, and for each one of the following choices, determine whether we have a legitimate sample space:

$\Omega = \{$Heads and it is raining, Heads and it is not raining, Tails$\}$

◯ Yes
◯ No

$\Omega = \{$Heads and it is raining, Tails and it is not raining, Tails$\}$

◯ Yes
◯ No

3.1.5 Sample space examples

Video: Sample space examples (transcripts)

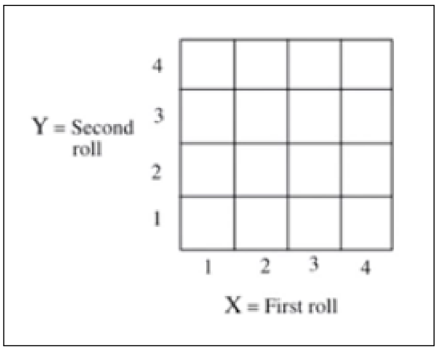
Let us now look at some examples of sample spaces. *Sample spaces are sets*. And a set can be:

- Discrete or continuous
- Finite or infinite

Let us start with a simpler case in which we have a sample space that is discrete and finite. The particular experiment we will be looking at is the following. We take a very special die, a tetrahedral die. So it's a die that has four faces numbered from 1 up 4. We roll it once. And then we roll it twice [again].

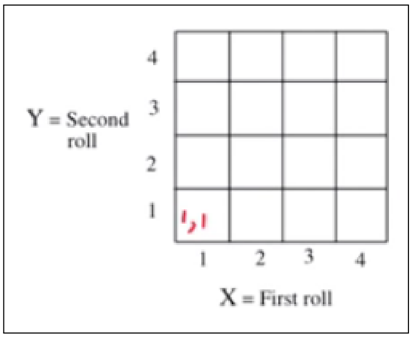
Were not dealing here with two probabilistic experiments. We're dealing with a single probabilistic experiment that involves two rolls of the die within that experiment. What is the sample space of that experiment?

Well, one possible representation is the following:

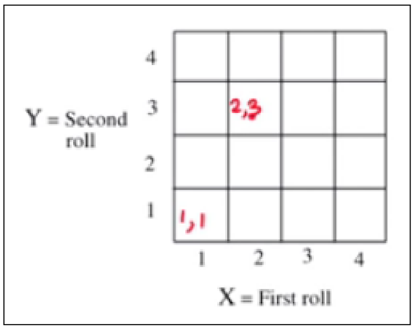


We take note of the result of the first roll. And then we take note of the result of the second roll. And this gives us a pair of numbers. Each one of the possible pairs of numbers corresponds to one of the little squares in this diagram.

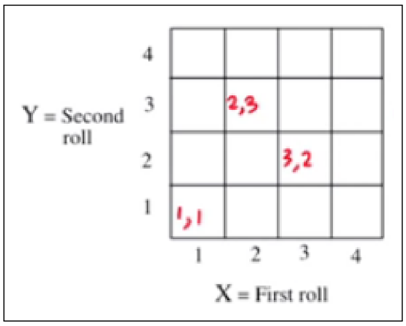
For example, if the first roll is 1 and the second is also 1, then this particular outcome has occurred:



If the first roll is it 2 and the second is a 3, then this particular outcome occurs:



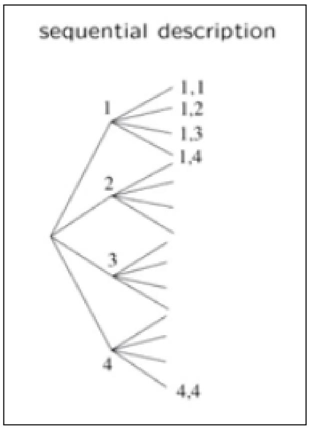
If the first roll is a 3 and then the next one is a 2, then this particular outcome occurs:



Notice that these two outcomes are pretty closely related. In both cases, we observe a 2 and we observe a 3. But we distinguish those two outcomes because in those two outcomes, the 2 and the 3 happen in different order. And the order in which they appear may be a detail which is of interest to us. And so we make this distinction in the sample space. So we keep the (3, 2) and the (2, 3) as separate outcomes.

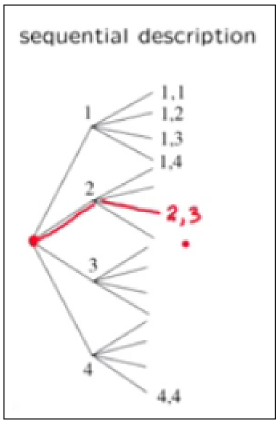
Now this is a case of a model in which *the probabilistic experiment can be described in phases or stages*. We could think about rolling the die once and then going ahead with the second roll. So we have two stages.

A very useful way of describing the sample space of experiments — whenever we have an experiment with several stages, either real stages or imagined stages — it is by providing a *sequential description in terms of a tree*.

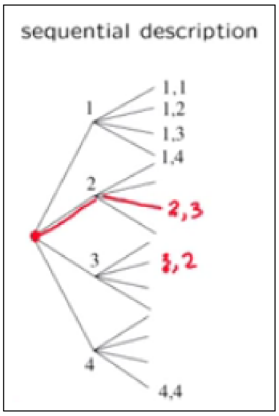


So a diagram of this kind, we call it a tree. You can think of this as the *root* of the tree from which you start. And the endpoints of the tree, we usually call them the *leaves*.

So the experiment starts. We carry out the first phase, which in this case is the first roll. And we see what happens. So maybe we get a 2 in the first roll. And then we take note of what happened in the second roll. And maybe the result was a 3. So we follow this branch here. And we end up at this particular leaf, which is the leaf associated with the outcome 2, 3:



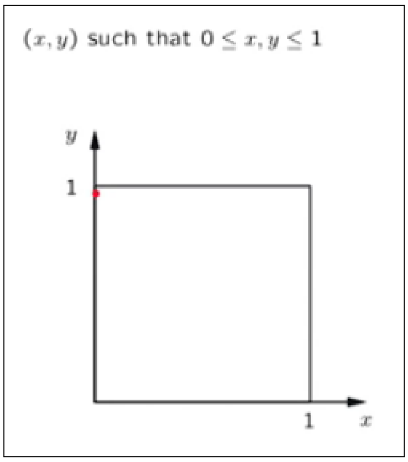
Notice that in this tree we once more have a distinction. The outcome 2 followed by a 3 is different from the outcome 3 followed by a 2, which would correspond to this particular place in the diagram:



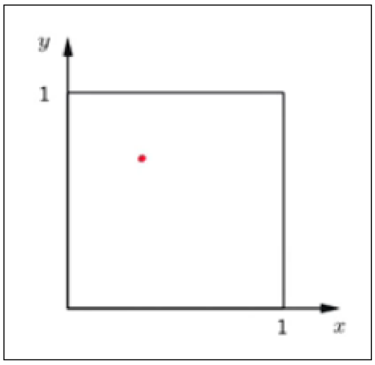
In both cases, we have 16 possible outcomes. 4 times 4 makes 16. And similarly, if you count here, the number of leaves is equal to 16.

The previous example involves a sample space that was discrete and finite. There were only 16 possible outcomes. But sample spaces can also be infinite. And they could also be continuous sets. Here's an example of an experiment that involves a continuous sample space.

So we have a rectangular target which is the unit square:



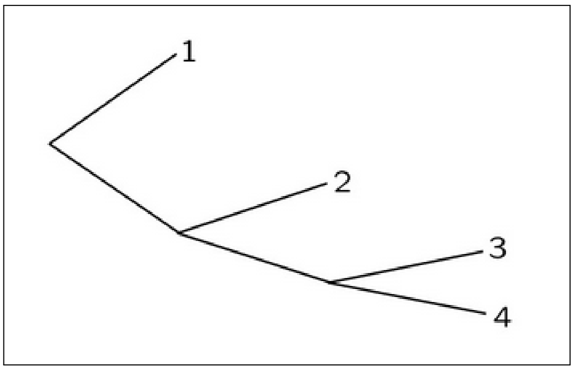
And you throw a dart on that target. And suppose that you are so skilled that no matter what, when you throw the dart, it always falls inside the target:



Once the dart hits the target, you record the coordinates x and y of the particular point that resulted from your dart throw. And we record x and y with *infinite precision*. So x and y are real numbers. So in this experiment, the sample space is just the set of x, y pairs that lie between 0 and 1 (inclusive): $\Omega = \{(x, y) \quad | \quad 0 \leq x, y \leq 1\}$

3.1.6 Exercise: Tree representations

Exercise 3.1.6-1: Tree representations
Paul checks the weather forecast. If the forecast is good, Paul will go out for a walk. If the forecast is bad, then Paul will either stay home or go out. If he goes out, he might either remember or forget his umbrella. In the tree diagram below, identify the leaf that corresponds to the event that the forecast is bad and Paul stays home.



- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4

3.1.7 Probability axioms

Video: Probability axioms (transcripts)

We have so far discussed the first step involved in the construction of a probabilistic model. Namely, the construction of a sample space, which is a description of the possible outcomes of a probabilistic experiment. We now come to the second and much more interesting part. We need to specify which outcomes are more likely to occur and which ones are less likely to occur and so on. And we will do that by assigning probabilities to the different outcomes. However, as we try to do this assignment, we run into some kind of difficulty, which is the following. Remember the previous experiment involving a continuous sample space, which was the unit square and in which we throw a dart at random and record the point that occurred. In this experiment, what do you think is the probability of a particular point? Let's say what is the probability that my dart hits exactly the center of this square. Well, this probability would be essentially 0. Hitting the center exactly with infinite precision should be 0. And so it's natural that in such a continuous model any individual point should have a 0 probability. For this reason instead of assigning probabilities to individual points, we will instead assign probabilities to whole sets, that is, to subsets of the sample space. So here we have our sample space, which is some abstract set omega. Here is a subset of the sample space. Call it capital A. We're going to assign a probability to that subset A, which we're going to denote with this notation, which we read as the probability of set A. So probabilities will be assigned to subsets. And these will not cause us difficulties in the continuous case because even though individual points would have 0 probability, if you ask me what are the odds that my dart falls in the upper half, let's say, of this diagram, then that should be a reasonable positive number. So even though individual outcomes may have 0 probabilities, sets of outcomes in general would be expected to have positive probabilities. So coming back, we're going to assign probabilities to the various subsets of the sample space. And here comes a piece of terminology, that a subset of the sample space is called an event. Why is it called an event? Because once we carry out the experiment and we observe the outcome of the experiment, either this outcome is inside the set A and in that case we say that event A has occurred, or the outcome falls outside the set A in which case we say that event A did not occur. Now we want to move on and describe certain rules. The rules of the game in probabilistic models, which are basically the rules that these probabilities should satisfy. They shouldn't be completely arbitrary. First, by convention, probabilities are always given in the range between 0 and 1. Intuitively, 0 probability means that we believe that something practically cannot happen. And probability of 1 means that we're practically certain that an event of interest is going to happen. So we want to specify rules of these kind for probabilities. These rules that any probabilistic model should satisfy are called the axioms of probability theory. And our first axiom is a nonnegativity axiom. Namely, probabilities will always be non-negative numbers. It's a reasonable rule. The second rule is that if the subset that we're looking at is actually not a subset but is the entire sample space omega, the probability of it should always be equal to 1. What does that mean? We know that the outcome is going to be an element of the sample space. This is the definition of the sample space. So we have absolute certainty that our outcome is going to be in omega. Or in different language we have absolute certainty that event omega is going to occur. And we capture this certainty by saying that the probability of event omega is equal to 1. These two axioms are pretty simple and very intuitive. The more interesting axiom is the next one that says something a little more complicated. Before we discuss that particular axiom, a quick reminder about set theoretic notation. If we have two sets, let's say a set A, and another set, another set B, we use this particular notation, which we read as "A intersection B"to refer to the collection of elements that belong to both A and B. So in this picture, the intersection of A and B is this shaded set. We use this notation, which we read as "A union B", to refer to the set of elements that belong to A or to B or to both. So in terms of this picture, the union of the two sets would be this blue set. After this reminder about set theoretic notation, now let us look at the form of the third axiom. What does it say? If we have two sets, two events, two subsets of the sample space, which are disjoint. So here's our sample space. And here are the two sets that are disjoint. In mathematical terms, two sets being disjoint means that their intersection has no elements. So their intersection is the empty set. And we use this symbol here to denote the empty set. So if the intersection of two sets is empty, then the probability that the outcome of the experiments falls in the union of A and B, that is, the probability that the outcome is here or there, is equal to the sum of the probabilities of these two sets. This is called the additivity axiom. So it says that we can add probabilities of different sets when those two sets are disjoint. In some sense we can think of probability as being one pound of some substance which is spread over our sample space and the probability of A is how much of that substance is sitting on top of a set A. So what this axiom is saying is that the total amount of that substance sitting on top of A and B is how much is sitting on top of A plus how much is sitting on top of B. And that is the case whenever the sets A and B are disjoint from each other. The additivity axiom needs to be refined a bit. We will talk about that a little later. Other than this refinement, these three axioms are the only requirements in order to have a legitimate probability model. At this point you may ask, shouldn't there be more requirements? Shouldn't we, for example, say that probabilities cannot be greater than 1? Yes and no. We do not want probabilities to be larger than 1, but we do not need to say it. As we will see in the next segment, such a requirement follows from what we have already said. And the same is true for several other natural properties of probabilities.