

## MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.431x: Probability—the Science of Uncertainty and Data

---

### WHAT TO DO AFTER THIS COURSE ?

---

Data science is a very important field these days. The term “data science” is fairly broad; it covers computer science issues such as how to collect, store, process, and share data, how to manage databases, etc. Probability theory can be used to model some of these computer systems but, more importantly, to process data and derive statistical conclusions. In fact, probabilistic analysis is central to a thorough understanding of most machine learning methods. On the other hand, basic probability theory by itself is certainly not enough for someone to be a successful practitioner of data science and machine learning.

Competence is generally achieved in two ways:

1. Learning more about the methodologies through courses, textbooks, etc. This may include courses on general statistical theory or machine learning, or courses geared to a particular field; for examples, economists tend to take courses in Econometrics.
2. Learning the **art** of statistical inference. How do I describe the data that I have collected? How do I come up with an initial model structure whose details I will try to estimate based on the available data? Which method should I use to carry out the estimation? What are meaningful conclusions that I can legitimately derive from the data, and what are some misinterpretations or pitfalls to avoid? As with every art, textbook learning is not enough - practice is needed, involving manipulation of real data and based on an understanding of the intricacies of a particular application domain. This is something that can be partly learned through project-based courses or through apprenticeship in either an academic or industrial setting.

## 1 What’s Next in the Micromasters in Statistics and Data Science (SDS) Program

The Micromasters in Statistics and Data Science program is a **multi-disciplinary program** with courses originating from different academic departments at MIT. Much like a multi-disciplinary program on campus, **the course structure, lecture style, effort requirement is different for each course**. We hope that the program as a whole will allow you to achieve competence in the ways described above, and at a level enough to get you started on your own unique data science journey, be it further studies, research, or applications to real life problems.

The two courses offered next term in the Micromasters in Statistics and Data Science (SDS) Program are two statistics courses that complement each other:

- [14.310x/Fx Data Analysis for Social Scientists](#)
- [18.6501x Fundamentals of Statistics](#)

## 1.1 14.310x/Fx Data Analysis for Social Scientists

The course *14.310x/Fx Data Science for Social Scientists* is **an introduction to statistics course** in which you will learn and gain intuition on common statistical methods. Rather than learning about the underlying theory, you will gain some skills and experience in applying these methods to real data sets, and learn about some accompanying practical issues.

This course **does not require 6.431x as prerequisite**. The first half will be on probability, which you know and love by now. We recommend that you use this opportunity to review probability notions from a slightly different perspective. The second half will be new material: confidence interval,  $p$ -values, hypothesis testing, linear regression and practical issues. Along the way, we will provide some guidance on R to prepare you for further exploration and research on your own.

You will learn from award winning MIT economist Professor Esther Duflo, whose research has helped change the way governments and aid organizations address global poverty, and MIT Lecturer Sara Fisher Ellison.

**Note:** Because this course is part of another MITx Micromasters (in Data, Economics, and Development Policy), the content of the course and the SDS Micromasters final exam are separated into two courses. To meet the requirement for the SDS Micromasters, **you must enroll and pass both the content course, *14.310x Data Analysis for Social Scientists*, and the final exam course, *14.310Fx Data Analysis in Social Sciences–Assessment*** (“F” stands for “final exam”). Please register for the content course at [14.310x](#); 14.310Fx will be open for enrollment within a month.

## 1.2 18.6501x Fundamentals of Statistics

This is a **mathematics course** in which you will dive into the mathematical underpinning behind the methods covered in 14.310x as well as more advanced statistical methods. You will expand your statistical knowledge to not only include a list of methods, but also the **mathematical principles that link these methods together, equipping you with the tools you need to develop new ones**.

This course **requires 6.431x, multivariable calculus, and knowledge on vectors and matrices as prerequisites**. You are expected to be comfortable with the probability concepts in 6.431x, as well as know how to compute gradients, Hessians (the higher-dimensional second derivatives), and write expressions in terms of matrices and vectors (e.g.  $M^T \sigma M$ , where  $M^T$  is the

transpose of the matrix  $M$ ). While we will not go into the proofs of the theorems, we will use mathematical formalism. We will not use software nor go into computation aspects.

You will learn from award-winning MIT Mathematics Professor Philippe Rigollet, who works at the intersection of statistics, machine learning, and optimization, focusing primarily on the design and analysis of statistical methods for high-dimensional problems.

If you would like to start preparing, this course will be a revised version of [OCW 18.650](#).

### 1.3 Recommendations and Time Commitment

We recommend that you **first take 14.310x/Fx, then 18.6501x**, but they can be taken concurrently and will complement each other.

Given that you are armed with probability, the course *14.310x/Fx Data Analysis for Social Scientists* **may require less time commitment than 6.431x** at roughly 8-14 hours a week. There will be weekly lecture exercises and problem sets, 1 final exam in the content course, and then the SDS Micro-masters final exam (14.310Fx) at the end.

The course *18.6501x Fundamentals of Statistics* will require **at least as much effort as 6.431x, and even more depending on your mathematics background**. There will be weekly lecture exercises and problem sets, 2 mid-term exams, and 1 final exam.

If you are able to commit more time than 14.310x/Fx takes especially in the beginning, it may be beneficial for you to start learning the material in 18.6501x at the same time. You can then decide to take it seriously for credit in the next run the term after next term. Keep in mind that 14.310x second half will require more effort.

## 2 Other courses at MIT

The rest of this document is intended to give you a sense of the academic side of the overall landscape. We will provide you with the descriptions of various courses that are available at MIT. (As a rule of thumb, any class with a zero following the dot (e.g., 6.011) is an undergraduate class. )

### 2.1 Systems and Signal Processing :

- **6.011 Introduction to Communication, Control, and Signal Processing:** ([OCW link](#))

Syllabus: This course examines signals, systems and inference as unifying themes in communication, control and signal processing. Topics include input-output and state-space models of linear systems driven by deterministic and random signals; time- and transform-domain representations in discrete and continuous time; group delay; state feedback and observers; probabilistic models; stochastic processes, correlation functions, power spectra, spectral factorization; least-mean square error estimation; Wiener filtering; hypothesis testing; detection; matched filters. s

- **6.341 Discrete-Time Signal Processing:** ([OCW link](#)) ([edX link](#))

Syllabus: This class addresses the representation, analysis, and design of discrete time signals and systems. The major concepts covered include: Discrete-time processing of continuous-time signals; decimation, interpolation, and sampling rate conversion; flowgraph structures for DT systems; time- and frequency-domain design techniques for recursive (IIR) and non-recursive (FIR) filters; linear prediction; discrete Fourier transform, FFT algorithm; short-time Fourier analysis and filter banks; multirate techniques; Hilbert transforms; Cepstral analysis and various applications.

### 2.2 Deeper into Theory:

- **6.436 Fundamentals of Probability:** ([OCW link](#))

Syllabus: This is a course on the fundamentals of probability geared towards first- or second-year graduate students who are interested in a rigorous development of the subject. The course covers most of the topics in 6.041 (sample space, random variables, expectations, transforms, Bernoulli and Poisson processes, finite Markov chains, limit theorems) but at a faster pace and in more depth. There are also a number of additional topics, such as language, terminology, and key results from measure theory; interchange of limits and expectations; multivariate Gaussian distributions; deeper understanding of conditional distributions and expectations.

- **Good references:** We list here a couple of great reference books and a set of reference videos. The books are much more advanced and would require a serious background in analysis.
  - **Probability: Theory and Examples**, *Rick Durrett* ([Amazon link](#))
  - **Probability with Martingales**, *David Williams* ([Amazon link](#))
  - **Probability Foundation**, *Krishna Jagannathan* ([Youtube channel link](#))

## 2.3 Statistics, Machine Learning, Inference:

- **6.008 Introduction to Inference:** ([OCW link](#)) ([class link](#))

Syllabus: Introduction to probabilistic modeling for problems of inference and machine learning from data, emphasizing analytical and computational aspects. Distributions, marginalization, conditioning, and structure; graphical representations. Belief propagation, decision making, classification, estimation, and prediction. Sampling methods and analysis. Introduction to asymptotic analysis and information measures, and applications.

- **18.443 Statistics for Applications:** ([OCW link](#))

Syllabus: This course offers a broad treatment of statistics, concentrating on specific statistical techniques used in science and industry. Topics include: hypothesis testing and estimation, confidence intervals, chi-square tests, nonparametric statistics, analysis of variance, regression, and correlation.

- **6.036 Introduction to Machine Learning:** ([class link](#))

Syllabus: Machine learning methods are commonly used across engineering and sciences, from computer systems to physics. Moreover, commercial sites such as search engines, recommender systems (e.g., Netflix, Amazon), advertisers, and financial institutions employ machine learning algorithms for content recommendation, predicting customer behavior, compliance, or risk. As a discipline, machine learning tries to design and understand computer programs that learn from experience for the purpose of prediction or control. In this course, you will learn about principles and algorithms for turning training data into effective automated predictions. We will cover concepts such as representation, over-fitting, regularization, and generalization; topics such as clustering, classification, recommender problems, and probabilistic modeling; and methods such as on-line algorithms, support vector machines, hidden Markov models, and Bayesian networks.

- **6.867 Introduction to Machine Learning:** ([OCW link](#))

Syllabus: 6.867 is an introductory course on machine learning which gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with topics such as classification and linear regression and ending up with more recent topics such as boosting, support vector machines, hidden Markov models, and Bayesian networks. The course will give the student the basic ideas and intuition behind modern machine learning methods as well as a bit more formal understanding of how, why, and when they work. The underlying theme in the course is statistical inference as it provides the foundation for most of the methods covered.

- **6.438 Algorithms for Inference:** ([OCW link](#))

Syllabus: This is a graduate-level introduction to the principles of statistical inference with probabilistic models defined using graphical representations. The material in this course constitutes a common foundation for work in machine learning, signal processing, artificial intelligence, computer vision, control, and communication. Ultimately, the subject is about teaching you contemporary approaches to, and perspectives on, problems of statistical inference.

- **Good reference:**

An excellent and concise introduction to all of the basic topics in statistics is:

- **All of Statistics**, *Larry Wasserman* ([Amazon link](#))

An excellent reference geared towards an understanding of modern statistical practice is:

- **An Introduction to Statistical Learning: with Applications in R**, *James et al.* ([Amazon link](#))

## 2.4 Random Processes:

- **18.445 Introduction to Stochastic Processes:** ([OCW link](#))

Syllabus: This course is an introduction to Markov chains, random walks, martingales, and Galton-Watson tree. The course requires basic knowledge in probability theory and linear algebra including conditional expectation and matrix.

- **6.262 Discrete Stochastic Processes:** ([OCW link](#))

Syllabus: Review of probability and laws of large numbers; Poisson counting process and renewal processes; Markov chains (including Markov decision theory), branching processes, birth-death processes, and semi-Markov processes; continuous-time Markov chains and reversibility; random walks, martingales, and large deviations; applications from queueing, communication, control, and operations research.

- **6.265 Advanced Stochastic Processes:** ([OCW link](#))

Syllabus: This class covers the analysis and modeling of stochastic processes. Topics include measure theoretic probability, martingales, filtration, and stopping theorems, elements of large deviations theory, Brownian motion and reflected Brownian motion, stochastic integration and Ito calculus and functional limit theorems. In addition, the class will go over some applications to finance theory, insurance, queueing and inventory models.

## 2.5 Decision Making in Uncertain Dynamic Environments:

- **6.231 Dynamic Programming and Stochastic Control:** ([OCW link](#))

Syllabus: The course covers the basic models and solution techniques for problems of sequential decision making under uncertainty (stochastic control). We consider optimal control of a dynamical system over both a finite and an infinite number of stages. This includes systems with finite or infinite state spaces, as well as perfectly or imperfectly observed systems. We also discuss approximation methods for problems involving large state spaces. Applications of dynamic programming in a variety of fields will be covered in recitations.

## 2.6 Applications:

Everywhere; look for them and enjoy! ☺ Having mastered the material in 6.431x, you already have the tools to deal with many of the applications that arise in different fields.