Probability and Statistics in Data Science using Python
Course Syllabus

Welcome!

We are delighted to welcome you to Probability and Statistics in Data Science using Python. In this course, you will learn the motivation, intuition, and theory behind the probabilistic and statistical foundations of data science, and will get to experiment and practice with these concepts via Python programs and the Jupyter Notebook platform.

Course Staff

Instructors

Alon Orlitsky, Professor, ECE and CSE Departments, UCSD

Yoav Freund, Professor, ECE Department, UCSD

Teaching Assistants

Mingchao Liang, Graduate Student, ECE, UCSD

Prerequisites

The course is intended for learners with an undergraduate degree, or senior undergraduates interested in broadening their understanding of probability and statistics. It assumes basic knowledge of the following topics

- Logic (e.g., De Morgan's Laws)
- Set theory (e.g., what are functions)
- Calculus (e.g., calculating integrals and derivatives)
- Programming (e.g., basic experience with any programming language)
- Linear algebra (e.g., vectors and matrices)

The Python programming language will be used throughout the course. This can be obtained by taking the course [DSE 200x, Python for Data Science](#).

Learning Objectives

The course will teach you how to visualize, understand, and reason about probabilistic and statistical concepts, and how to apply your knowledge to analyze data sets and draw meaningful conclusions from data. We will cover both theoretical and practical aspects, and will start each topic with motivation and intuition and will proceed with rigorous arguments and provable techniques. Each topic will be accompanied by a Python Notebook that you could run and modify to experiment with the material learned and get a better feel for the material covered.

Topics

The course will cover twelve topics:
1. Overview of sets
2. Counting principles
3. Combinatorics
4. Discrete probability
5. Conditional probability and Bayes' Rule
6. Random variables, expectation, variance, and correlation
7. Common distribution families
8. Continuous distributions
9. Probabilistic inequalities, concentration, and limit theorems
10. Regression
11. Sampling, parameter estimation, and confidence intervals
12. Hypothesis testing

Python Notebooks

Each topic will be accompanied by a Python Jupyter Notebook with programs for visualizing, simulating, and exploring the material covered. You will be able to run the programs, modify the algorithms, and experiment with the simulations to get a better feel and understanding of the concepts covered.

Discussion forums

Two discussion forums, one for verified-learners and one for audit-learners, will let you to better understand the material by discussing it with fellow students and with course staff. While both forums will be monitored by our staff, the verified-learner forum will be more closely monitored by the course team.

Assignments and Exams

To help you better understand the material and help us assess your progress we will have 12 assignments, one associated with each topic, and a final exam.

Assignments (60% of grade)

Each of the 12 topics will consist of 3 components (2 of them graded):

- **Poll Questions (0% of grade - optional).** Simple polls are available for your benefit throughout the course, they will not be graded.
- **Problem sets (35% of grade).** Each lecture will be followed by a small problem set of roughly two math or calculation problems that test your understanding of the fundamental and practical implications of the material studied. There will be roughly 70 such problem sets, of which we will drop the worst 14 (roughly 20%).
- **Programming** and **Data assignments (25% of grade).** One Python question to help you develop your programming skills and ability to implement probabilistic concepts. You will need to run the program on your computer and enter your numerical answer on EdX. We will drop the worst 2 (20%) grades.

● **Final Exam (40% of grade).**
The final exam will cover all course materials. It will consist of 10 problems similar in nature to the problem sets and several problems may be taken or modified from the homework assignments. Due to testing challenges, programming problems will not be included in the final. A practice exam will be posted to prepare you for the type of questions asked. You can bring clean scratch paper(s), one standard sheet of paper (two sides, front and back) of handwritten or typed notes, pens/pencils, and a physical calculator. **Note: no other material will be allowed.** You will have up to 4.5 hours to complete the exam.

Time and grading policies for weekly assignments
   ● **Worst grades dropped.** For two of the graded components (problem sets and programming questions), we will automatically drop for you 20% of your worst grades. This means for example that you can skip two assignments (topics) for the programming problems, 14 problem set assignments and still receive full marks.

Verified Learner - Earning a Certificate
To earn a verified certificate for this course, you need to enroll as part of the verified track, complete identity verification, and earn a passing grade. If you are auditing the course, you will not receive a certificate.

Becoming A Verified Learner
Your deadline to switch from the Audit Learner track to the Verified Learner track can be found on the Course Home Page.

Grading
Letter grades will be assigned as follows. Grades of "B" and above are considered passing grades.

Percent of total points	Grade
85%-100%	A
65%-84.9%	B
<65%	F

Effort
We expect the course to take 10-15 hours per week to complete.

Pace and deadlines

The course is self-paced, all materials will stay online until the end of the course. As described above, you will have until the end of the course to submit all graded assignment.

Honor Code
Beyond learning this important material, we hope that you will take the course seriously and respect your fellow students. Please read and abide by the [edX honor code pledge](). If you have any questions about what is permitted or not please let us know.

We value your feedback
This online course is new for us and for UC San Diego. We tried hard to provide you with high-quality education, but there will certainly be room for improvement. If there is anything, large or small, we can improve, please do let us know. We greatly appreciate your feedback and will take it very seriously.

Thank you!
Thank you very much for taking the course. We hope you will find it educational and useful.