Statistics and Probability in Data Science using Python
Course Overview

Welcome!

We are delighted to welcome you to Statistics and Probability 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

Matthew Elliott, Graduate Student, CSE, UCSD
Rohit Parasnis, Graduate Student, ECE, UCSD
Hanwen Yao, Graduate Student, ECE, UCSD
Zhen Zhai, Graduate Student, CSE, UCSD

Prerequisites

The course is intended for learners with an undergraduate degree or senior undergraduates interested in broadening their understanding of probability and statistics. We will assume 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. If you would like to learn or gain more practice with Python, please consider viewing or taking the first course in this MicroMasters, 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 the following topics:
- Counting and combinatorics
- Discrete and continuous probability
- Conditional probability and Bayes' Rule
- Random variables
- Expectation, variance, and correlation
- Common distribution families
- Probabilistic inequalities and concentration
- Moments and limit theorems
- Hypothesis testing
- Sampling and confidence intervals
- PCA and regression
- Entropy and compression

Course Outline
The course consists of 10 units. In each of the course's first 10 weeks we will release one unit, and you will have six weeks to complete it.
- Week 1 - Introduction
- Week 2 - Sets
- Week 3 - Counting and combinatorics
- Week 4 - Probability and conditioning
- Week 5 - Random Variables, expectation, and variance
- Week 6 - Discrete and Continuous distribution families
- Week 7 - Inequalities and concentration theorems
- Week 8 - Sampling, confidence intervals, and hypothesis testing
- Week 9 - Regression and Principal Component Analysis
- Week 10 - Entropy and Compression

Python Notebooks
Each week we will post a Python Jupyter Notebook with programs for simulating, visualizing, and testing the material covered. You will be able to run the programs and modify them to experiment with the algorithms and get a better feel the concepts covered.

Discussion forums
Discussion forums will enable students better understand the material by discussing it with each other and with course staff. While both forums will be monitored by our staff, the verified-learner forum will receive priority in time and answers.

Assignments and Exams

To help you better understand the material and help us assess your progress we will have 10 weekly assignments, and a final exam.

● **Weekly assignments (65% of grade)**

Each assignment will consist of 5 components (4 of them graded):

- ● **Engagement (5% of grade).** Meant to encourage you to watch the videos and view the associated materials, all you need to do is check the "mark as complete" button.
- ● **Poll Questions (0% of grade - optional).** Simple polls are available for your benefit throughout the course, they will not be graded.
- ● **Comprehension Quizzes (10% of grade).** About 10 simple multiple-choice questions based on the week's videos.
- ● **Problem sets (25% of grade).** About 10 math or calculation problems that test your understanding of the fundamental and practical implications of the material studied.
- ● **Programming Assignments (25% of grade).** A Python assignment to help you develop your programming skills and ability to implement probabilistic concepts.

● **Final exam (35% of grade).**

The final exam will cover the full course material. We are checking with EdX about the exam's timing and format and will post the information as soon as we have it.

Time and grading policies for weekly assignments

- ● **Six-week assignment due time**. Each of the 10 assignments will be due within 6 weeks from the time becomes available. For example the first assignment will be available at the beginning of week one and you can submit it anytime till the end of week 6, while the last, 10th, assignment will be available at the beginning of week 10 and can be submitted till the end of week 15, when the course ends. You can submit each component at a different time.
- ● **Worst 2 grades dropped.** For each of the four weekly graded components (engagement, comprehension quizzes, problem sets, and programming assignments), we will automatically drop for you the two worst weekly grades. This means for example that you can skip two weeks for the programming assignment, and two other weeks for the quizzes, and still receive full score.

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. You can change from unverified to verified track until the end of the sixth week.

Grading

Letter grades will be assigned as follows. Grades C and above are considered passing grades.

Percent of total points	Grade

85%-100%	A
70%-84.9%	B
50%-69.9%	C
<49.9%	F

Effort
We expect the course to take 8-10 hours per week to complete.

Pace and deadlines
The course is instructor-paced, and every week we will release the relevant material for that week. We encourage you to keep current with the material released, however, all materials will stay online till the end of the course. As described above, you will have six weeks to submit each weekly 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.