Course: STAT W4700
Title: Probability and Statistics for Data Science
Semester: Fall 2014

1 Course Description

Objective: Statistics is the language in which data is analyzed and interpreted, and thus any serious data scientist must have a firm understanding of the mathematical principles of probability and statistics. Assiduous students of this course will build this critical foundation.

Details: This course is a self-contained introduction to probability and statistics with a focus on data science. The topics covered include fundamentals of probability theory and statistical inference, including: probabilistic models, random variables, useful distributions, expectations, the law of large numbers, the central limit theorem, point and confidence interval estimation, maximum likelihood methods, hypothesis tests, and linear regression (as time permits).

Prerequisites: Current familiarity and facility with basic calculus; some background in calculus up to multivariate integration; a background in basic optimization up to the use of Lagrange multipliers.

Note: This syllabus will change throughout the term. In particular, the homework and the schedule will be adjusted per class pace.

2 Course Logistics

Time: Tu/Th, 6:10pm-7:25pm
Location: Mathematics 520
Instructor:

John P. Cunningham
Email: `jpc2181@columbia.edu`
Office: Department of Statistics 1026, 1255 Amsterdam Ave

Teaching Assistant:

Haolei Weng
Email: `hw2375@columbia.edu`
Office: Department of Statistics 1025, 1255 Amsterdam Ave
Traditional Office Hours: Mondays, 6:30-8:30pm

Virtual Office Hours:

Owing to student schedules and the subsequent challenges of finding mutually suitable office hours, we will use a virtual platform. Piazza is a highly regarded forum for students to discuss class questions, homework problems, and more. Discussing problems is encouraged, but full solutions should not be posted (see section on academic integrity). The tool can be found at: `https://piazza.com/class/hz9yj3aja4r1mx`. Many Columbia classes find this to be a much quicker and more effective tool than traditional office hours, and we strongly encourage students to use it both to ask questions and to improve their own understanding by posting answers and comments.

Textbook (Required):

Probability and Statistics, Fourth Edition
M. H. DeGroot and M. J. Schervish
Publisher: Pearson (2012)
ISBN: (ISBN-10) 0321500466; (ISBN-13) 978-0321500465
Note: This textbook is available at Book Culture.
Note: Many homework problems will be drawn from this text, so the correct edition is required.

3 Grading and Academic Integrity

We take the honor code very seriously; students caught cheating or otherwise in violation will face disciplinary action. Please note the Barnard honor code text:

> "We... resolve to uphold the honor of the College by refraining from every form of dishonesty in our academic life. We consider it dishonest to ask for, give, or receive help in examinations or quizzes, to use any papers or books not authorized by the instructor in examinations, or to present oral work or written work which is not entirely our own, unless otherwise approved by the instructor.... We pledge to do all that is in our power to create a spirit of honesty and honor for its own sake."
> `http://barnard.edu/node/2875`
> `https://www.college.columbia.edu/academics/academicintegrity`

Grading: Grading will be on a curve and weighted according to:

Weekly Homework (lowest 2 are automatically dropped)	30%
Midterm Exam	30%
Final Exam	40%

Late Work and Regrading Policy: No late work or requests for regrades are accepted.

Homework: Students are encouraged to work together, but homework write-ups must be done individually and must be entirely the author's own work. Homework is due at the **beginning** of each Tuesday class, the week after it is assigned. To receive full credit, students must thoroughly explain how they arrived at their solutions and include the following information on their homeworks: name, UNI, homework number (e.g., HW03), class (STAT W4700). Homework must be turned in online through Courseworks, must be in PDF format, have a .pdf extension (lowercase!), and be less than 4MB. Homeworks not adhering to these requirements will receive no credit.

Midterm and Final Exams: Students may bring one sheet of 8.5×11in paper, on which the student can write notes (front and back). No other materials are allowed (including textbooks, computers, and other electronics). To receive full credit, students must thoroughly explain how they arrived at their solutions.

4 Schedule and Reading

Week	Date	Content (subject to change)
1	9/2 - 9/7	• Topics: introduction, sample spaces, axioms, combinatorics • Reading: Ch 1.1-1.11 (skim: 1.1-1.2; skip ❖ in 1.4 only) • HW01: 1.4.10, 1.5.4, 1.6.8, 1.7.8, 1.8.12, 1.9.8, 1.10.2
2	9/8 - 9/14	• Topics: conditional probability, independence, Bayes rule, random variables • Reading: Ch 2.1-2.4; Ch 3.1 • HW02: 2.1.6, 2.1.14, 2.2.10, 2.2.23, 2.3.4, 2.3.6, 2.5.28
3	9/15 - 9/21	• Topics: random variables, distributions, cdf's, marginals, conditionals • Reading: Ch 3.1-3.6 (intentional repetition of 3.1) • HW03: 3.1.8, 3.2.4, 3.3.4, 3.3.8, 3.4.4, 3.4.10, 3.5.2, 3.6.12
4	9/22 - 9/28	• Topics: multivariate distributions, functions of a r.v., expectation • Reading: Ch 3.6-3.9 (particularly 3.7) • HW04: 3.4.6, 3.5.10, 3.6.10, 3.7.8, 3.8.8, 3.9.4 (use Jacobian)
5	9/29 - 10/5	• Topics: moments, covariance • Reading: Ch 4.1-4.3; 4.5-4.7 • HW05: 4.1.8, 4.2.4, 4.2.6, 4.3.6, 4.5.6, 4.6.14, 4.7.6 • Note: No Lecture Thursday 10/02; self-review
6	10/6 - 10/12	• Topics: useful distributions and their special properties, review • Reading: Ch 5.1-5.8 (skip ❖ in 5.4), 5.10 • HW06: 5.2.6, 5.2.10, 5.4.6, 5.6.8, 5.6.10, 5.7.6, 5.7.8, 5.11.8, 5.11.24 • Note: HW06 due in two weeks (Tuesday 10/21)
7	10/13 - 10/19	• Topics: bivariate normal distribution, review • Reading: (continued from last week) • HW06: (continued from last week) • Note: **MIDTERM EXAM** Thursday Oct 16
8	10/20 - 10/26	• Topics: law of large numbers, central limit theorem • Reading: Ch 6.1-6.4 (skip all ❖ in 6.2 and 6.3) • HW07: 6.1.2, 6.2.5, 6.2.8, 6.3.4, 6.3.8, 6.3.10, 6.5.9
9	10/27 - 11/2	• Topics: math review, maximum likelihood • Reading: Ch 7.1-7.2; 7.5-7.7 • HW08: 7.5.2, 7.5.10, 7.6.4
		continued on next page...

...continued from previous page

Week	Date	Content (subject to change)
10	11/3 - 11/9	• Topics: maximum likelihood, sampling distributions • Reading: Ch 8.1-8.7 • HW09: 8.1.2, 8.3.6, 8.4.2, 8.4.6, 8.5.2, 8.5.4, 8.6.2, 8.7.2 • Note: HW09 due in two weeks (Tuesday 11/18)
11	11/10 - 11/16	• Topics: sampling distributions • Reading: (continued from last week) • HW09: (continued from last week)
12	11/17 - 11/23	• Topics: hypothesis testing, p-values, t-tests, • Reading: Ch 9.1-9.2; 9.5 • HW10: 9.1.2, 9.1.4, 9.2.6, 9.5.2, 9.5.4, 9.5.12, 9.5.14 • Note: HW10 due in two weeks (Tuesday 12/2)
13	11/24 - 11/30	• Topics: (continued from last week) • Reading: (continued from last week) • HW10: (continued from last week) • Note: Thursday 11/27 class cancelled (Thanksgiving)
14	12/1 - 12/7	• Topics: p-values, t-tests, two sample t-tests, F test, χ^2 test, review • Reading: Ch 9.5-9.7; 10.1 • HW11: 4.6.8, 6.2.7, 7.5.12, 9.5.16, 9.6.6, 9.7.4, 9.7.7, 10.1.4
15	12/8 - 12/19	• **FINAL EXAM**