USC Viterbi School of Engineering

DSCI 564: Probability and Statistics for Data Science (Fall 2025)

Units: 4

Instructor: Mohammad Reza Rajati, PhD
GCS 302-B
rajati@usc.edu – Include DSCI 564 in subject

Office Hours: Right after the lecture, by appointment

Webpage: Personal Homepage at Intelligent Decision Analysis

TA(s): Will be introduced on Piazza

Lecture(s): Tuesday, Thursday, 4:00 - 5:50 pm, DMC 150 & Online

Webpages: Piazza Class Page for everything except grades
and USC Brightspace Class Page for grades and homework submission
– All HWs, handouts, solutions will be posted in PDF format
– *Student has the responsibility to stay current with webpage material*

Prerequisites: Prior courses in multivariate calculus, linear algebra,
and linear system theory.

Other Requirements: Basic computer skills (e.g., plotting, Matlab, Excel, Python, etc.).
Use of R is mandatory.
Therefore, the students need to know R or be willing to learn R.

Tentative Grading: Assignments 15%
Three Midterm Exams 75%
Final Project 10%
Participation on Piazza* 5%

Letter Grade Distribution:

≥ 93.00	A	73.00 - 76.99	C
90.00 - 92.99	A-	70.00 - 72.99	C-
87.00 - 89.99	B+	67.00 - 69.99	D+
83.00 - 86.99	B	63.00 - 66.99	D
80.00 - 82.99	B-	60.00 - 62.99	D-
77.00 - 79.99	C+	≤ 59.99	F

Disclaimer: Although the instructor does not expect this syllabus to drastically change, he reserves every right to change this syllabus any time in the semester.

Note on e-mail vs. Piazza: If you have a question about the material or logistics of the class and wish to ask it electronically, please post it on the piazza page (not e-mail). Often times, if one student has a question/comment, other also have a similar question/comment. Private Piazza

posts should be used to contact the professor, TA, graders only for issues that are specific to you individually (e.g., a scheduling issue or grade issue).

Catalogue Description: This course introduces fundamental concepts in probability and statistics from a data science perspective. It aims at synergistically presenting rigorous probabilistic reasoning and problem solving as well as computerage statistical methods that are widely used in data science.

Course Objectives: Upon successful completion of this course a student will

- Understand axiomatic probability and know how to model real-world problems using it
- Understand Discrete and Continuous Random Variables, their distributions, their properties, moments, and correlations.
- Understand the limiting behavior of large amounts of data by limit theorems
- Understand sampling and sampling distributions
- Cunstruct models of distributions using histograms and density estimation techniques
- Estimate parameters of distributions using maximum Likelihood, maximum a-posteriori, and other estimation techniques
- Assess the properties of estimators
- Construct confidence intervals for point estimates
- Test hypotheses about different parameters of distributions of populations using samples of data
- Construct linear regression models for data, assess those models, and select variables using various techniques including statistical tests and regularization
- Use non-parametric and robust statistical tools to assess classification and regression models
- Apply resampling and Monte-Carlo methods to computation and statistical inference
- Use statistical software (R, SPSS, STATA, Python, etc.) to effectively use all of the techniques learned in the course
- Be ready for understanding and implementing machine learning and data mining methods that rely on statistical analysis.

Exam Dates:

- **Midterm Exam 1 (in person):** Thursday, Oct 2, 4:00-5:50 PM, in class.
- **Midterm Exam 2 (in person):** Thursday, Nov 6, 4:00-5:50 PM, in class.
- **Midterm Exam 3 (in person):** Thursday, Dec 4, 4:00-5:50 PM, in class.

- **Final Project Due:** Tuesday, December 11, 6:00 PM as set by the university. **Grace period**: the project can be submitted until 11:59 PM of the same day with 30% penalty. Any change in the project after the deadline is considered late submission. One second late is late. The project is graded based on *when it was submitted, not when it was finished.* Homework late days *cannot* be used for the project.

Textbooks:

- **Required Textbooks:**

1. *Probability and Random Processes for Electrical and Computer Engineers*, 1st Edition
   Author: John A. Gubner; Cambridge University Press, 2006. **ISBN-13:** 978-0521864701
2. *Probability with Applications and R*, 1st Edition
   Author: Robert P. Dobrow; Wiley, 2014. **ISBN-13:** 978-1-118-24125-7

- **Recommended Textbooks:**

1. *Probability and Random Processes*, 3rd Edition
   Authors: Geoffery R. Grimmet and David R. Stirzaker; Oxford University Press; 2001. **ISBN-13:** 978-0198572220
2. *Introduction to Probability*, 2nd Edition
   Authors: Dimitri P. Bertsekas and John N. Tsitsiklis; Athena Scientific, 2008. **ISBN-13:** 978-1886529236
3. *Introduction to Probability Models*, 11th Edition
   Authors: Sheldon M. Ross, Academic Press, 2010. **ISBN-13:** 978-0124079489
4. *One Thousand Exercises in Probability*, 1st Edition
   Authors: Geoffery R. Grimmet and David R. Stirzaker; Oxford University Press; 2001. **ISBN-13:** 978-0198572213
5. *Schaum's Outline of Probability, Random Variables, and Random Processes*, 3rd Edition
   Author: Hwei P. Hsu; McGraw-Hill Education; 2014. **ISBN-13:** 978-0071368100
6. *Schaum's Outline of Probability and Statistics*, 4th Edition
   Authors: John J. Schiller Jr., R. Alu Srinivasan, Murray R Spiegel; McGraw-Hill Education; 2012. **ISBN-13:** 978-0071795579
7. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, 1st Edition
   Authors: Bradly Efron and Trevor Hastie; Cambridge University Press, 2016. **ISBN-13:** 978-1107149892
8. *Probability and Statistics for Data Science*, 1st Edition **Author:** Norman Matloff; Chapman and Hall, 2019. **ISBN-13:** 978-1138393295
9. *Statistical Inference*, 2nd Edition
   Authors: George Casella and Roger L. Berger; Duxbury, 2001. **ISBN-13:** 978-0534243128
10. *An Introduction to Statistical Inference and Its Applications with R*, 1st Edition
    Author: Michael W. Trosset; CRC Press, 2009. **ISBN-13:** 978-1584889472

11. *Introduction to Mathematical Statistics*, 8th Edition
    Authors: Robert V. Hogg, Joseph W. McKean, and Allen T. Craig; Pearson, 2018. **ISBN-13:** 978-0-13-468699-8
12. *Mathematical Statistics with Resampling and R*, 2nd Edition
    Authors: Laura M. Chihara and Tim C. Hesterberg; Wiley, 2019. **ISBN-13:** 978-1119416548

Grading Policies:

- The letter grade distribution table guarantees the *minimum* grade each student will receive based on their final score. When appropriate, relative performance measures will be used to assign the final grade, at the discretion of the instructor.
  - Final grades are non-negotiable and are assigned at the discretion of the instructor. If you cannot accept this condition, you should not enroll in this course.
  - Three of your lowest homework grades will be dropped from the final grade.
  - Half of your lowest midterm grade will be dropped. For example, if you receive 70, 60 and 40 in three midterms, your total midterm grade will be $\frac{70+60+0.5\times40}{2.5} = 60$ instead of $\frac{70+60+40}{3} = 56.7$.
  - *Participation on Piazza has up to 5% extra credit, which is granted on a competetive basis *at the discretion of the instructor*.

- **Homework Policy**
  - Homework is assigned on an approximately weekly basis. Homework due dates are mentioned in the course outline, so mark your calendars. A three-day grace period can be used for each homework with 10% penalty per day. Any change in homework after the deadline makes it a late submission. *Absolutely no late homework will be accepted after the grace period. A late assignment results in a zero grade.*
  - Late Days: No late homework will be accepted after the three day grace period. One second after the deadline is considered late. However, students are allowed to use *six* late days for homework for any reason (including sickness, family emergencies, overwhelming workload, exams, etc) *without incurring the 10% penalty.* Beyond that, no individual extension will be granted to anyone for any reason whatsoever.

    Example: A student can submit six assignments, one day late each, without any penalty. Or three assignments, two days late each, without penalty, or two assignments three days late each. A student cannot use four late days for one assignment, and two late days for another assignment. An assignment submitted four days late will receive a zero grade, although its grade will be dropped as the lowest homework grade, according to the above grading policies.
  - Use your six late days strategically and only if you absolutely need them. Always remember that later in the semester, you might become sick or have heavy workload in other courses and might need to use your late days.
  - Poor internet connection, failing to upload properly, or similar issues are NOT acceptable reasons for late submissions. If you want to make sure that you do not have such problems, submit homework eight hours earlier than the deadline. Please do not ask the instructor to make individual exceptions.

  - Homework is graded based on *when it was submitted, not when it was finished.*
  - Homework solutions should be typed or *scanned* using scanners or mobile scanner applications like CamScanner and uploaded on the course website (photos taken by cell-phone cameras and in formats other than pdf will NOT be accepted). Programs and simulation results have to be uploaded on the course website as well.
  - Students are encouraged to discuss homework problems with one another, but each student must do their own work and submit individual solutions written/ coded in their own hand. Copying the solutions or submitting identical homework sets is written evidence of cheating. The penalty ranges from F on the homework or exam, to an F in the course, to recommended expulsion.
  - Posting the homework assignments and their solutions to online forums or sharing them with other students is strictly prohibited and infringes the copyright of the instructor. Instances will be reported to USC officials as academic dishonesty for disciplinary action.

- **Exam Policy**

  - **Make-up Exams:** No make-up exams will be given. If you cannot make the above dates due to a class schedule conflict or personal matter, you must drop the class. In the case of a required business trip or a medical emergency, a signed letter from your manager or physician has to be submitted. This letter must include the contact of your physician or manager.

    Important Note: If you have emergencies, you should state them before taking the exam. Taking the exam, waiting for the grade, and then mentioning that you were sick *is not be acceptable*
  - Exams will be closed book and notes. No calculators are allowed nor are computers and cell-phones or any devices that have internet capability. One letter size cheat sheet (back and front) is allowed for the first two midterms. Two letter size cheat sheets (back and front) are allowed for the third midterm.
  - All exams are cumulative, with considerable emphasis on material presented since the last exam.

- **Attendance:**

  - Students are required to attend all the lectures and discussion sessions and actively participate in class discussions. Use of cellphones and laptops is prohibited in the classroom. If you need your electronic devices to take notes, you should discuss with the instructor at the beginning of the semester.

Important Notes:

- Textbooks are secondary to the lecture notes and homework assignments.
- Handouts and course material will be distributed.
- Please use your USC email to register on Piazza and to contact the instructor and TAs.

Tentative Course Outline

Tuesday	Thursday
Aug 26th **1** **Introduction** **Logic**	28th **2** **Set Theory**
Sep 2nd **3** **Set Theory** , **Probability Models** • Sample Space, • σ-algebra of events • Probability as An Additive Measure • Continuity of Probability • Conditional Probability	4th **4** **Probability Models and Independence** • Total Probability • The Baye's Rule • The Multiplication Rule
9th **5** **Random Variables** • Definitions • CDFs • Borel Sets	11th **6** **Random Variables** • CDFs • Independence • Multiple Random Variables **Combinatorics (Reading)**

Tuesday	Thursday
16th **7** **Discrete Random Variables** PMFs Famous Discrete Random Variables	18th **8** **Discrete Random Variables** • Famous Discrete Random Variables • Multiple Random Variables • Joint PMFs • Marginal PMFs • Conditional PMFs • Total Probability • Substitution Law • Independence • Derived Distributions
23rd **9** **Moments of Discrete Random Variables** • Expectation • The Law of The Unconscious Statistician • Properties of Expectation • Higher Order Moments • Variance and Standard Deviation	25th **10** **Moments of Discrete Random Variables** • Moments of Famous Discrete Random Variables • Existence of Expectations* • Covariance and Correlation and Their Properties • Expectation As Norm and Inner Product • The Cauchy-Schwartz-Bunyakovsky Lemma

Tuesday	Thursday
30th **11** **Moments of Discrete Random Variables** • Expectation As Norm and Inner Product • The Cauchy-Schwartz-Bunyakovsky Lemma **Conditional Expectation** • The Law of The Unconscious Statistician • Substitution Law for Conditional Expectation • Total Expectation	Oct 2nd **12** **Conditional Expectation** • Conditional Expectation as A Random Variable • Properties of Conditional Expectation • Existence of Conditional Expectation • Conditional Probability as Conditional Expectation • Wald's Equality • Projections, Projection Theorem, Principle of Orthogonality • Conditional Expectation as an Estimator
7th **13** **Continuous Random Variables** • PDFs • Important Continuous Random Variables	9th Fall Recess
14th **14** **Continuous Random Variables** • Important Continuous Random Variables • Multiple Random Variables and Joint PDFs • Marginal PDFs • Independence • Conditional Probability and Conditional PDFs • Moments of Continuous Random Variables	16th **15** **Continuous Random Variables** • Existence and Properties of Moments • Moments of Famous Continuous Random Variables • The Law of The Unconscious Statistician (LOTUS)

Tuesday	Thursday
21st **16** **Continuous Random Variables** • The Law of Total Probability • The Substitution Law • Total Probability • Total Expectation • Total Probability and Expectation for Multiple Random Variables • Conditional Expectation **The Bivariate Normal Distribution**	23rd **17** **Random Vectors** • Expectation of A Random Vector • Linearity of Expectation • Auto-correlation Matrix • Covariance Matrix • Positive Definiteness • Cross-correlation Matrix • Cross-covariance Matrix • The Multivariate Normal Distribution **Derived Distributions** • Monotonic Functions • Linear Functions
28th **18** **Derived Distributions** • Non-Monotonic Functions • Multivariable Functions • Linear Mappings • A Single Function of Multiple Random Variables • Order Statistics	30th **19** **Derived Distributions** • Order Statistics • Sum of Independent Random Variables • Noraml Random Variables in Polar Cooridnates • The Rayleigh Distribution • Simulation of Random Variables • The Box-Muller Method • Rejection Sampling Algorithm

Tuesday	Thursday
Nov 4th **20** **Generating Functions** • Moment Generating Functions • Region of Convergence • Inversion of MGFs • Properties of MGFs	6th **21** **Generating Functions** • Random Sums of Random Variables • Laplace and Z transforms • Characteristic Functions • Generating Functions for Random Vectors • Joint Characteristic Functions
11th Veterans Day	13th **22** **Concentration Inequalities** • Markov and Chebychev Inequalities **Stochastic Convergence** • Modes of Convergence • Hierarchy of Modes of Convergence
18th **23** **Limit Theorems** • Weak Law of Large Numbers • Strong Law of Large Numbers • Monte-Carlo Methods • Bootsrtap* • The Central Limit Theorem • Berry-Esseen Theorem • Binomial Approximation • Chi-squared Approximation	20th **24** **Statistics** • Histograms • Kernel Density Estimation • Point and Interval Estimation of The Mean • One-Sided and Two-Sided Confidence Intervals • Interpretation of Confidence Intervals • Estimation of Variance • Student's T-Statistic

Tuesday	Thursday
25th **25** **Statistics** • Point and Interval Estimation of Proportion • Two Sample Confidence Intervals • Interval Estimation of Difference between Means (Independent and Dependent Samples) • Interval Estimation of Ratio of Variances • The Fisher-Snedecor Statistic • Boostrap Confidence Intervals* • Frequentist (Fisherian) Hypothesis Testing • p-values* • Type-I and Type-II Errors • Power of A Test • Neyman-Pearson Lemma* • Testing for The Mean, Proportion, Difference in The Means, and Difference in The Proportions • The Kolmogorov-Smirnov Test • The Chi-Squared Test	27th Thanksgiving Break

Tuesday	Thursday
Dec 2nd **26** **Statistics** • Parameter Estimation • Properties of Estimators • Method of Moments • Minimum Variance Unbiased Estimator* • Maximum Likelihood Estimation • The Cramér-Rao Bound • Maximum A-Posteriori Estimate • Minimum Mean-Squared Error Estimate	4th **27** **Statistics: Linear Regression*** • Simple Linear Regression • Multiple Regression • Least Squares • Confidence Intervals and Hypothesis Testing for Coefficients • Multicollinearity • Heteroscedasticity • F-test for ANOVA and Overal Significance of Model

Notes:

- Items marked by * will be covered only if time permits.
- Instead of Markov Chains, Robust Statistical Methods may be covered.

Homework Due Dates & Exams

Monday	
Aug 25th -	**1**
Sep 1st -	**2**
8th **Homework 1 Due (Moved to Tuesday Sep 5)**	**3**
15th **Homework 2 Due**	**4**
22nd **Homework 3 Due**	**5**
29th **Homework 4 Due**	**6**
Oct 6th **Homework 5 Due**	**7**
13th **Homework 6 Due**	**8**
20th **Homework 7 Due**	**9**
27th **Homework 8 Due**	**10**
Nov 3rd **Homework 9 Due**	**11**
10th **Homework 10 Due**	**12**

Monday	
17th **Homework 11 Due**	**13**
24th **Homework 12 Due**	**14**
Dec 1st **Homework 13 Due**	**15**

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on Research and Scholarship Misconduct.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
eeotix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
osas.usc.edu
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 821-4710
campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu,emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
ombuds.usc.edu
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-3340 or otfp@med.usc.edu
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.