

CS A263: PROBABILITY AND STATISTICS FOR COMPUTER SCIENCE

Item	Value
Curriculum Committee Approval Date	02/10/2021
Top Code	070600 - Computer Science (Transfer)
Units	4 Total Units
Hours	90 Total Hours (Lecture Hours 63; Lab Hours 27)
Total Outside of Class Hours	0
Course Credit Status	Credit: Degree Applicable (D)
Material Fee	No
Basic Skills	Not Basic Skills (N)
Repeatable	No
Grading Policy	Standard Letter (S), • Pass/No Pass (B)

Course Description

Introduction to probability and statistics with an emphasis on their applications in Computer Science. Topics include continuous and discrete probability distributions, linear and logistic regression, creating models to use for predictive inference, and programmatic analysis of data. PREREQUISITE: CS A131, CS A150, or CS A170; and MATH A180. Transfer Credit: CSU; UC.

Course Level Student Learning Outcome(s)

1. Students will apply appropriate distributions to compute the probability of a specific event or specific events.
2. Students will apply counting techniques to compute the probability of an event.
3. Students will define a linear regression model to describe a dataset and use the model to infer the likelihood of specific events with respect to that dataset.

Course Objectives

- 1. Define and use a probability distribution to compute the probability of a specific event.
- 2. Define a random variable and its distribution to compute the probability of specific event(s).
- 3. Compute the mean, standard deviation, and confidence intervals of a given data set.
- 4. Create a Linear Regression model to describe the relationship between linearly related variables.
- 5. Compute the expected value and variance of a random variable.
- 6. Write a program using a language like R or Python to compute basic statistics about a dataset.
- 7. Use a language such as R or Python to create visualizations to analyze correlations between variables in a data set.
- 8. Use linear regression as a means of inference to solve simple machine learning tasks.

- 9. Use logistic regression as a means of inference to solve simple machine learning tasks.
- 10. Use a language such as R or Python to analyze a data set to state whether a hypothesis about the data is probable, such as confirming when specific weather will most likely occur.
- 11. Evaluate claims made by media sources by analyzing publicly available data sets and determining if the statistical descriptors accurately reflect the data as a whole.
- 12. Explain why variance in data affects a computational models accuracy.

Lecture Content

Probability Fundamentals Sample Spaces Permutations and Combinations Conditional Probability Bayes Theorem Independence and Dependence Variance and Covariance Probability Distributions Discrete Distributions Poisson Distribution Binomial Distribution Geometric Distribution Continuous Distributions Normal Distributions Exponential Distribution Gamma Distribution Chi-Squared Distribution Point Estimation Density Functions Random Variables Discrete and Continuous Random Variables Distributions over Random Variables Expected Value Variance Descriptive Statistics Mean Standard Deviation Confidence Intervals Central Limit Theorem Linear Regression Maximum Likelihood Estimation Using regression for inference Using linear regression for predicting missing data Using linear regression for predicting future events Logistic Regression Maximum Likelihood Estimation Using logistic regression for predicting missing data Using logistic regression for predicting future events Explore how accuracy is limited by a computers ability to represent floating point numbers and how using logistic models helps to solve some of these problems Using Software to Analyze Data Clean and visualize data sets Create and analyze graphs of data with respect to correlation amongst data variables Compute the mean, standard deviation, and confidence intervals programmatically Computing the probability of events based on large amounts of data Using Software for Inference Create linear regression models to infer unknown gaps in data Create logistical regression models to infer unknown gaps in data Compare different models used for inference with respect to their accuracy Explaining why variance in the data affects the accuracy of a given model

Lab Content

The following is a list of possible programming labs that are related to the content topics: Data visualization Create graphs of different variables on opposing axes to analyze possible correlations. Visualize different probability distributions using different parameters (such as mean or variance). Empirical probability Model various probability distributions to find the probability of a specific point or event. Descriptive statistics Programatically compute the mean, standard deviation, confidence levels, etc of a dataset. Correlation Analyze different dataset variables for possible correlations between them. Verifying or Refuting Media Claims Using a publically available dataset, determine if the claim made about the data is statistically sound (such as claims about weather patterns, shopping patterns, housing prices, economic shifts, etc). Models of Inference Create a linear regression model to predict missing data. Create a logistical regression model to predict missing data. Using a regression model to predict future events.

Method(s) of Instruction

- Lecture (02)
- DE Live Online Lecture (02S)

- DE Online Lecture (02X)
- Lab (04)
- DE Live Online Lab (04S)
- DE Online Lab (04X)

Instructional Techniques

Lecture, demonstration, and in-class exercises.

Reading Assignments

Students will spend a minimum of 4 hours per week reading the textbook. Students will be expected to follow along with the exercises in the reading material.

Writing Assignments

Students will spend a minimum of 3 hours per week writing code.

Out-of-class Assignments

Students will spend a minimum of 6 hours per week completing weekly problem set assignments.

Demonstration of Critical Thinking

Students will demonstrate the ability to solve a variety of statistical problems with respect to computer science or data analysis.

Required Writing, Problem Solving, Skills Demonstration

Students will demonstrate proficiency in applying statistical software to analyze data.

Eligible Disciplines

Computer science: Masters degree in computer science or computer engineering OR bachelors degree in either of the above AND masters degree in mathematics, cybernetics, business administration, accounting or engineering OR bachelors degree in engineering AND masters degree in cybernetics, engineering mathematics, or business administration OR bachelors degree in mathematics AND masters degree in cybernetics, engineering mathematics, or business administration OR bachelors degree in any of the above AND a masters degree in information science, computer information systems, or information systems OR the equivalent. Note: Courses in the use of computer programs for application to a particular discipline may be classified, for the minimum qualification purposes, under the discipline of the application. Masters degree required.

Textbooks Resources

1. Required Diez, D., Barr C., Cetinkaya-Rundel, M.. OpenIntro Statistics, 3 ed. OpenIntro, 2015 2. Required Devore, J. Probability Statistics for Engineering and the Sciences, ed. Cengage Learning, 2016