Course Outline
Semester 1 2021/22

Course Title COMPUTATIONAL STATISTICS I
Course code: STAT 6181
Credits: 3
Lecturer Dr Isaac Dialsingh
Contact hours: Mondays (5-8pm) - 3 hours per week
Email address: isaac.dialsingh@sta.uwi.edu
Pre-requisites: Some Knowledge of statistical inference

Course Description and Rationale

Research in the Computational Statistics and Simulation group involves modeling, analyzing, and simulating dynamic systems characterized by complex process logic and uncertain behaviors. These tools are applied in a variety of areas, such as computational finance, environmental monitoring, object/target recognition, statistical signal and image processing, crime analysis, health care delivery, logistics and distribution, manufacturing, and aerospace systems.

The use of computers in statistics is becoming increasingly popular. This course is presented to address these concerns. Students will be mainly the has three main goals:

(1) for students to learn about computationally intensive approaches to probability and statistical inference, including Monte Carlo, Markov chain Monte Carlo methods, the bootstrap, the expectation-maximization (EM) algorithm, some matrix analysis, and some select topics (time permitting);

(2) using computational tools to reinforce important concepts from probability and statistics;

(3) to develop fluency with statistical computing using the language R.

The main topics covered in the course:

• Random number generation, Monte Carlo basics.
• Importance sampling, Markov chain Monte Carlo.
• Parametric and nonparametric bootstrap, permutation tests.
• Optimization basics. Expectation-maximization (EM) algorithm.
- Bayesian Inference. An introduction.

Content

Aims and Goals

Computational Statistics is a branch of mathematical sciences concerned with efficient methods for obtaining numerical solutions to statistically formulated problems. This course will introduce students to a variety of computationally intensive statistical techniques and the role of computation as a tool of discovery.
Topics include numerical optimization in statistical inference (Expectation Maximization (EM) algorithm, Fisher scoring, etc.), random number generation,
Monte Carlo methods, randomization methods, jackknife methods, bootstrap (parametric and non-parametric) methods,
Introduce and understand modern computational methods used in statistics. Included are methods for simulation, estimation and visualization of statistical data.
Understand the role of computation as a tool of discovery in data analysis.

Be able to appropriately apply computational methodologies to real world statistical problems.

Objectives:
Upon successful completion of this course, students MUST be able to:
- State and use appropriate optimization techniques for single and multi-parameter problems
- Use bootstrap in both the parametric and non-parametric setting
- Use the Jack-knife in parameter estimation
- Find the Expectation and Maximization steps of the EM Algorithm given a statistical scenario.
- Generate random numbers from various distributions
- Use R to solve problems involving numerical approximations

Mode of Delivery
Lectures delivered face-to-face. All lectures, assignments, handouts, and review materials are **available online** to all students. Lectures are supplemented with laboratory work.

Course content and structure (Tentative)

1 Introduction to Course and other Preliminaries
Review of the main numerical recipes in R. An introduction to matrix algebra using R. Writing short routines in R.
Course introduction, format of delivery

2 Optimization Theory
Single and Multivariate methods

3 EM Algorithm Review of the Expectation and Maximization. Steps and theory

4 EM Algorithm Use of software to solve problems such as finding the number of components in a mixture of normal.

5 Sampling from discrete and continuous distributions

6 Basic Monte Carlo Integrals and Importance Sampling

7 Metropolis Hastings algorithm Lectures, with lecture notes made available

8 Bootstrap
The Basics and Application to tests and confidence intervals

9 Bootstrap
Advanced topics – Other bootstrap methods

10 Permutation and The Jacknife Its advantages and disadvantages
Applications to microarray data

11 Bayesian Statistics

12 Group Presentations

Assessment Course-work 100 %
This course will be assessed completely via 4 individual assignments and one group project. Each assignment and project will involve both theoretical and computer based problems. This is a tentative breakdown.

Course work Examinations (2) – 30%
There would be two in course examinations.
Individual Assignments (3) – 40%
Three homework assignments will be given, collected and graded throughout the semester. These assignments must be typed and the Rcode must be submitted as email attachments for grading.
While discussion of the homework is allowed, you must prepare your solutions separately. Direct copying of written work or computer code is considered cheating and will result in a zero on the assignment. Assignments are worth 30% of the course grade.

Group Project (1) – 30%
The minimum group size is 3, however larger groups are encouraged. The topics will vary and can be discussed with the instructor. The groups will be required to present their project in class on the last week of classes. Full details will be given around class session four. The project is worth 30% of the course grade.

Resource requirements
The statistical computing necessary will be R software.

PRESCRIBED TEXTS AND READING MATERIALS

Required reading
Computational Statistics, by G. H. Givens and J. A. Hoeting, (Wiley 2005).
Statistical Computing with R by M. Rizzo, Chapman and Hall
(This book was available last year on the UWI Net library system free of charge).
Recommended reading
Hastie, T., Tibshirani, R. and Friedman J. 2009. Elements of Statistical Learning Springer.