STAT6181
Computational Statistics I
Lecture 1

Course Overview

Overview

- The course specifics
- The content
- The assessments
- The Textbooks
- R Software

The Course Specifics

- Level: Graduate Level Semester: I
- No. of Credits: 3
- Lecturer: Dr Isaac Dialsingh
- Email: isaac.dialsingh@sta.uwi.edu
- Telephone: 662-2002 Ext 83554
-
- **Pre-requisites:**
- Some knowledge of Statistical Inference is beneficial.
- **Co-requisites** None

Why this course is important?

- Computational Power is cheap

- Assumptions in Statistics do not hold all the time.

When this happens, we often have some computational methods available

- Not all problems have exact solutions. In cases where an exact solution does not exist, we can “estimate” the answer.

- Research today especially with the prevalence of big data makes computational methods necessary.

The Content

- 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 - the basics.
  - Imputation Techniques
  - Expectation-maximization (EM) algorithm.
  - Bayesian Inference.

Random number generation - Monte Carlo basics.

- Generating random values from common distributions
  - common univariate distributions like exponential
  - uncommon univariate distributions (like the one below)
  - multivariate

x	1	2	4
P(X=x)	**0.2**	**0.3**	**0.5**

Random number generation, Monte Carlo basics.

- Multivariate distributions (discrete case)

	Y		
X	**0**	**1**	**2**
0	0.1	0.1	0.2
1	0.3	0.2	0.1

Estimation of Integrals

- Can you integrate this?

$$\int_{-\infty}^{\infty} 100e^{-(x-3)^2} \, \mathrm{dx}$$

OR even this

$$\int_{0}^{\infty} x^4 e^{-2x} \, \mathrm{dx}$$

Parametric and nonparametric bootstrap, permutation tests.

- In inferential statistics, we often make assumptions about the distribution of data, for example, we may assume normality which most times lead to normality of a particular test statistic.

- But what happens when the normality assumption is in question?

- This is where the bootstrapping can assist us.

Simple example : Bootstrap

Suppose we have a sample of size 10 from a population that is not normal. How can we find a 95% confidence interval for the population mean?

Optimization Techniques

- There are simple optimization problems and off course there are difficult ones.
- Estimation problems in Statistics are essentially optimization problems.
- We will meet MLEs, Least Squares Method, all of which utilizes optimization techniques
- Some MLE problems are ‘exact’ while others need to be estimated - the problem usually boils down to the solution of a system of equations after the differentiation etc.

Some distributions

Two parameter Distributions

- Normal distribution EASY!!!

- Gamma distribution? HARD?

Imputation Techniques (if time permits)

How do we handle missing data?

- Delete the rows of data missing?
- Impute the missing values with a number(s)? Which numbers should we use?
- Perform multiple imputation.

The EM (Expectation Maximization) Algorithm

Estimate parameters using a “missing” data approach

Works well when no analytic solution is present.

Mixture model example:

Collect data: 12, 10, 11, 14, 12, 10, 4, 2, 4, 1, 3

How can we model the data? If we can model the data, can we estimate the parameters?

Bayesian Statistics

What it is?

How do we model data?

Assessments (tentative)

Coursework: 100%

- Group Project: 30%
- 2 Take home Tests (30%)
- 3 Individual Assignments (40%)

The Textbooks

- **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
- **Recommended reading**
- Hastie, T., Tibshirani, R. and Friedman J. 2009. Elements of Statistical Learning Springer.

R Software

Found everywhere in this course!!!

Questions?