STAT 341 - Computational Statistics and Data Analysis

Cameron Roopnarine

Last updated: January 12, 2021

Contents

Chapter 1

Introduction

1.1 Preamble

- The subject matter of *computational* statistics is that of *Statistics* itself, but developed via computation rather than only *through mathematics*.

  Essentially, we are rebuilding our understanding of statistics through a computational perspective to be more general.

- The goal of the course is to present essential statistical concepts.
  - Simulation is used to illustrate the concepts and to provide understanding.
  - Mathematical development provides an alternative presentation of the same ideas, when that is possible, and is used to develop a tool or get insight into a concept.
- Because simulation is the primary means to develop this understanding.
  - The statistics/estimators/tests/etc. used should be of sufficient complexity that a complete mathematical treatment would be beyond the level of this course.
- Because the statistics/estimators/tests/etc. can be complex, several numerical methods are introduced. For example: Gradient descent, Newton's method, iteratively re-weighted squares, etc.

1.1.1 Programming in R

- Students are expected to program in R; other languages will not be accepted.
- The ideas is to convey programming concepts alongside the statistical concepts. The purpose of the programming, like mathematics, is to illustrate the statistical concepts.
  - Therefore, we will take advantage of the functional programming language R to write some general purpose code.
- The clarity and simplicity of the code is primary since it is being used to convey the statistical concepts; that is, the code is not production-level code, but instead teaching-level code.
- Wh