Johns Hopkins Engineering

625.464 Computational Statistics

Why Study Computational Statistics?

Module 1 Lecture 1A

Why Study Computational Statistics?

Statistics has been defined as:
  The transformation of raw data into knowledge.

E. Wegman 1988

Questions:

  What data should be collected?
  How much data should be collected?
  What conclusions can we draw from
  the data?
  How for can those conclusions be trusted?

What is Computational Statistics?

Computational Statistics is a collection of techniques that have a strong focus on the exploitation of computing in the creation of new statistical methodology.

Made possible by inexpensive computing hardware.

Data collected without any clear idea of what it will be used for.

Inspires creation of new statistical methodologies.

How does Computational Statistics Compare to Traditional Statistics?

Traditional Statistics

1. Small to moderate sample size
2. IID data sets
3. One or low dimensional
4. Manual computations
5. Mathematically tractable
6. Well focused question
7. Strong unverifiable assumptions
8. Statistical inference
9. Mainly closed form algorithms
10. Statistical optimality

Computational Statistics

1. Large to very large sample size
2. Nonhomogeneous data sets
3. High dimensional
4. Computationally intensive
5. Numerically tractable
6. Imprecise Question
7. Weak or no assumptions
8. Structural inference
9. Iterative algorithms possible
10. Statistical robustness

Basically....

Traditional Statistics: First design the study based on a research question, then collect the required data.

Computational Statistics: First collect the data, then design a study to gain some useful information from it.

We view Computational Statistics as a tool of discovery.