Post-Strike Recap

Max Turgeon

STAT 3150–Statistical Computing

Changes to the assessment schedule i

- We have 3 assignments and 2 exams left.
- I'll provide flexibility: let me know if you need extensions.
- I also don't expect you to work during the Holiday break.

Assessment	New Date
Assignment 4	Dec 17
Midterm 2	Dec 21
Assignment 5	Jan 19
Assignment 6	Jan 19

Changes to the assessment schedule ii

- Some comments:
  - Midterm 2 will be 80 minutes, but you can take it anytime during a 24h period.
  - The midterm will cover the material from before the strike.
  - Assignments 5 and 6 will be equivalent to a single assignment, but they will cover different topics.
  - A new final exam schedule will be provided soon.

Lectures in December

- I understand that some of you will be writing exams in the next few weeks.
    - And attending lectures will be difficult.
- I've uploaded all the remaining material to UM Learn, **including pre-recorded video lectures**.
    - They're all ~30 minutes, so shorter than a typical Zoom recording.
- I will also upload Assignments 5 and 6 next week.

In other words, you could finish the course on your own before Christmas; or you could focus only on Assignment 4 and the Midterm and catch up in January; or anything in between. It's up to you.

Recap

Main theme

- Recall the main theme of the course: **using computational techniques to solve statistical problems**.
- What kind of statistical problems?
    - Point estimation
    - Interval estimation
    - Hypothesis testing

Numerical methods and Optimisation

- For the first two modules, we specifically looked at point estimation.
- We talked about the following methods:
    - Bisection/Brent's method for root finding in one dimension.
    - Newton-Raphson for optimisation in any dimension.

Generating random variates

- R has many built-in functions for generating random variates.
    - `runif`, `rnorm`, etc.
- We discussed general techniques when these functions aren't enough.
    - Inverse transform, or generally any type of transformation.
    - Accept-reject sampling.
- **When would you need to generate random variates?**
    - Estimate expected values (i.e. Monte Carlo integration)
    - Estimate probability statements
    - Simulation studies

Monte Carlo integration i

- This topic mostly falls under *point estimation*.
- Estimate quantities of the form

$$E(g(X)) = \int g(x)f(x)dx, \qquad X \sim f.$$

- Trace plot = diagnose convergence issues
- Variance reduction
  - Antithetic variables
  - Control variates
  - Importance sampling
- Confidence intervals in MC integration are based on the Central Limit Theorem

Monte Carlo integration ii

- Since our estimates are sample means, we need to divide by $\sqrt{n}$, where n is the number of variates in the sample mean.

- **When would you use MC integration?**
  - To estimate difficult integrals.
  - Many, many estimators can be defined as expected values of transformations $g(X)$ of a random variable X.

Importance sampling

- It's a form of **variance reduction** for Monte Carlo integration.
- Based on the following identity:

$$E_f(g(X)) = E_\phi\left(\frac{g(X)f(X)}{\phi(X)}\right),$$

  as long as ϕ is nonzero on the support of f.
- We want to choose the importance function ϕ such that:
  - ϕ is a density from which it is "easy" to sample.
  - the ratio $\frac{|g(X)|f(X)}{\phi(X)}$ is almost constant.
- **Why do we care so much about reducing variance anyway?**
  - Because smaller variance means smaller confidence intervals, which means more accurate inference.

Monte Carlo methods for Inference

- This module was an interlude, connecting Monte Carlo integration and resampling methods.
    - What is a statistic? An estimator? A sampling distribution?
    - What is a type I error? Type II error? Power?
- If we are willing to completely specify the data generating mechanism, we can study the consequences of these assumptions through **Monte Carlo simulation**.
    - Which estimator is more efficient (i.e. has smallest variance)?
    - Does my confidence interval have the right coverage probability?
    - Which hypothesis test has largest power?