Stat 405/705
Class 10
Statistical computing with R

Richard P. Waterman

Wharton

Table of contents I

Today's module

Topics to be covered in this module:

- Last time
- Simulation modeling
- Case study: level of effort
- Functions used in today's class
- Next time

Last time

- Random variable generation
- Discrete random variables
- Continuous random variables
- Mixtures of random variables
- Multivariate random variables
- Cumulative Distribution Function
- Quantiles

Case study background

- People who own copyrighted *works* don't usually like them being infringed
- The owners are often represented by industry organizations
- These organizations attempt to deter infringement
- One approach is to monitor computer networks, identify infringers and directly deter them
- This requires cooperation from Internet Service Providers (ISPs)
- Deterrence costs money
- So implementation becomes a question about required resources to achieve a desired goal

The Graduated Notice Program

- One mode of deterrence is a Graduated Notice Program
- See: https://en.wikipedia.org/wiki/Graduated_response
- Three strikes:
  1. Education
  2. Acknowledgement
  3. Action/remediation

Framing the question

- The level of effort: how many notices do we have to send out a week, in order to reduce the number of infringers over the course of a year by 25%?
- Important facts to bear in mind
  1. Not every infringer gets a notice when infringing (it's random as to who gets caught) as scanning resources are limited
  2. Not every infringer has the same chance of getting caught
  3. Whether someone responds to a notice is not guaranteed
  4. Even if you send an infringer a notice, there is no guarantee that they receive it

Setting up the problem

Plan: set up a virtual world that tracks the effectiveness of the notice program.
There are three states we need to track at the infringer/week level

- Is a user still infringing? [Z]
- Did the user receive a notice? [Y]
- Was the user sent a notice? [X]

Random elements

- Being identified (caught/captured) as infringing in any week (can vary by person)
- Whether a sent notice is received
- Responding to a notice if you get it (may vary with notice level)
- Initial assumptions:
  - All infringers have the same probability of being identified (uniform)
  - More realistic: this probability varies across users, some are easier to catch than others
  - All notices (1, 2, and 3) have the same probability of success
  - More realistic: some notices have more value than others

Simulation components

Initially fixed components:

- The initial number of infringers: `n.infringers = 1000`
- The number of weeks in the program: `n.weeks = 52`
- The deterrence probabilities: `theta = c(0.1,0.25,0.1)`

What we will vary:

- The level of effort: notices sent per week, `effort = seq(10,50,5))`

Initial assumptions

- No migration. That is, no new infringers come into the infringer pool. None exit the pool (give up infringing by choice).
- Infringers are identified independently of one another.
- Once someone stops infringing, they don't start again.
- All of these could be relaxed within the simulation framework.

The loop structure

```
For each level of effort:
    For each week:
        See who is still infringing [Z]
        Send notices [X]
        See who gets them [Y]
        See who responds to the notice by stopping infringing
        Summarize current infringers
    Next week
    Summarize remaining week 52 infringers and level of effort
Next level of effort
```

Look at the source code

```
#### Simulation case study
set.seed(19390909) # Set the seed for reproducibility
# and let's see how it works
```

Look at the source code

Additional activities with the simulation

- Save results to .csv files
- Create graphics during the simulation
- Concatenate graphics to make an animation
- Save simulation parameters for future reference
- Consider parallelizing the outer loop

Module summary

Topics covered today include:

- Monte Carlo case study

Next time

- The R eco-system and add-on packages/libraries