STAT40180 — Stochastic Models

Brendan Murphy

Week 1

Models

Models

- We now consider potential models for each of the motivating examples:
  - Crime Statistics
  - Taxi Cab Problem
  - Employee Retention
  - Motorcycle Crash
  - Space Shuttle

Crime Statistics

- Let's assume that:
  - the data from each week are independent.
  - the number of crimes happening per week is constant over the data collection period.
  - the probability of a crime being reported is the same for all weeks and crimes.
- What model does this suggest?
- It suggests that the data can be modeled by a Binomial(n, p), where n and p are both unknown.

Taxi Cab Problem

- Let's assume that:
  - the taxis are numbered consecutively.
  - the taxi number doesn't affect it being observed outside the airport.
- What model does this suggest?
- We can assume that each number observed is a draw from a uniform distribution on the numbers $1, 2, \ldots, N$ where N is the unknown number of taxis in the city.

Employee Retention

- Let's assume that:
  - the employees retention times are independent.
  - the times are non-negative.
- What model does this suggest?
- We could use any probability distribution which accommodates positive values:
  - exponential
  - gamma
  - Weibull
  - log-normal
- We would need to allow the parameters to depend on the covariates, if these are available.

Motorcycle Crash

- The relationship between acceleration and time is clear, but it is complex.
- Standard linear regression models won't fit very well.
- If we could change the regression assumption from

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

to

$$Y_i = s(x_i) + \epsilon_i,$$

where $s(\cdot)$ is a "smooth" function, then we may be able to better model the crash.

Space Shuttle

- Let's assume that:
  - the launches are independent.
  - the number of failed O-rings is

$$\text{Binomial}(6, p),$$

    where p depends on the launch temperature.
- What model does this suggest?
- We could fit a binomial regression model:

$$Y_i \sim \text{Binomial}(6, p(x_i))$$

  where

$$p(x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

  .

Famous Quote

- For each scenario being modeled, we made a series of assumptions.
- This allowed us to posit a stochastic model for the scenario.
- It could be argued that some of the assumptions are unrealistic.
- However, we may still gain useful information from the modeling excercise.
- George Box once said,

  All models are wrong, but some models are useful