

Test 1 Info Sheet

Stat 341 – Spring 2017

Logistics

- In-class test: Friday, March 3
- Take-home test: Due Monday, March 6
- Both portions cover through chapter 5 of *Statistical Rethinking*

Topics

This list isn't meant to be exhaustive, but I hope it is useful in preparing for the test.

Techniques

1. Creating Numerical and Graphical Summaries
 - You have your choice of `lattice`, `ggplot2`, or `ggformula`, but I'm likely going to be leaning heavily and `ggformula`, since it does what we need most easily. If you use `lattice`, the migration should be pretty easy. If you already know (or are learning) `ggplot2`, it's fine to continue with that. (Anything you can do with `ggformula` can be done with `ggplot2`, since `ggformula` just translates a new interface into `ggplot2` code.
2. Grid Method
3. Quadratic approximation using `map()`
4. Posterior sampling
 - a. manually (for example, when using the grid method)
 - b. using `extract.samples()`
5. Using posterior samples to investigate
 - a. model parameters
 - b. model predictions (for averages and for individuals)
 - Creating counterfactual data (for example, for use with `link()` and `sim()`)
6. Centering
7. A non-exhaustive list of R commands you should know
 - Plotting commands (`lattice`, `ggplot2`, or `ggformula` versions) for the types of plots we have been making.
 - from `rethinking`: `map()`, `alist()`, `precis()`, `extract.samples()`, `link()`, `sim()`, `chainmode()`
 - `apply()`
 - working with data frames: `data_frame()`, `expand.grid()`, `mutate()`, `filter()`
 - distributions: `dnorm()`, `dbinom()`, `dunif()`, `rnorm()`, `rbinom()`, `runif()`
 - miscellaneous: `zscore()`, `str()`, `head()`, `sample()`, `coerce_index()`

Concepts

1. Conditional probability and connections to Bayesian inference
 - definition of conditional probability
 - problems like the pandas problem
2. $\text{posterior} \propto \text{prior} \cdot \text{likelihood}$
3. What grid apprid approximation is, how it works, and why it is limited to models with only a few parameters.
4. What posterior sampling is and what can be done with posterior samples.
5. Why quadratic approximation is called that and roughly how it works.
 - potential problems with quadratic approximation (and why we will need a third method eventually)
6. Creating models
 - selecting priors that are reasonable
 - creating a model relationship (equation with variables and parameters)
 - (additive) linear models
 - handling categorical variables
7. Interpreting models
 - interpreting model coefficients (and their uncertainty).
 - interpreting model predictions (and their uncertainty).
 - interpreting σ in models where the response is modeled with a normal distribution
 - plots for models (especially those with multiple predictors)
 - residuals
 - masking, spurious relationships and other issues that can be revealed by fitting models with multiple predictors.
8. Why and how to center variables used in models