Review homework and lecture slides

Review models closely

Review homework 5 closely

**Probability:**

Random experiment: outcome, sample sace, event, probability measure, iid

Probability model: a random exp. That mimics data generation while making assumption that you can disprove

**Prob. Examples 1:**

If necleotides in S are iid and uniform, what is P(S = ACGAA)?

.25 ^ 5

(Markov Chain)

.25 \* .2 \* .8 \* .2 \* .6 = 3/125

**Statistics:**

Sample (simple random sample)

Population (set of similar items or events)

Statistical inference: parameter estimation, hypothesis testing

Random variable: a real number computed on outcome of random experiment (a function that operates on outcomes)

Statistic: a rela number computed from the data (a function that operates on data)

p-value: the probability of data as or more extreme than the observed data (also, prob. We are wrong in concluding H0 is not correct)

vocabulary: siginificance level, bias, variance, type I error, type II error

Type I Error: Rejecting H0 when it is true

Type II Error: Inferring H0 when H0 is false

**Prob. Example 2:**

Collect seq data and count ACGT.

You are given outcome of experiment

Sample space: collection of all sets of counts for ACGT in permissive/nonper. Cells

What must be true of the data for relative G/C content to be higher in the permissive cells?

(NpG + NpC) / NpTotal must be > (NnG + NnC) / NnTotal

What is probability of events mentioned in 3 if Nn = Np and nucleotides are iid multinoulli(.25, .25, .25, .25)?

½

How can you estimate the prob. Using simulation?

1. N ~ poisson(alpha)
2. For I = 1 … n
3. SI ~ multinoulli(.25 … .25)
4. Count GC

No questions with t-Test or z-Test

**Python Questions:**

Benefits of programming instead of manual

Pros/cons of python

Two numeric types in python

Whats wrong with: score = 87/100 (python casts as int, so it becomes 0)

Can’t start variable names with number

**Machine Learning:**

Different types:

Supervised (classification (used in hw) and regression)

Unsupervised

Clustering

Reinforcement

Semi-supervised