1 Introduction
2 Context/Background information
- 2.1 overview of goodness of fit and different testing crietrions (one dimensional)
  - 2.1.1 analysis of how well each can differentiate between two samples from same distribution with small differences
  - 2.1.2 comparison of % of failed p values in exact sample

3 Requirements and analysis
- 3.1 problem statement
- 3.2 list of inference problems

4 design and implementation
- 4.1 overview of the website
- 4.2 overall structure of the code base
- 4.3 weights
  - 4.3.1 how different one dimensional tests were adjusted for weighted case
- 4.4 p value
  - 4.4.1 how p values are evaluated in one/two sample ks test
  - 4.4.2 how a single p value is extended to samples of p value (and percentage of which that has passed the test)
    - 4.4.2.1 how the initial sample is divided into either overlapping and non overlapping samples and its correlation
    - 4.4.2.2 (lack of) correlation between sample size and % passed
- 4.5 basic sampling algorithms
  - 4.5.1 how importance sampling, mcmc, rejection algorithms were designed and how they work
  - 4.5.2 their relative performance with respect to distribution of p values (and % of p values above critical value)
  - 4.5.3 their relative performance in estimating the mean (time and first sample size to get the difference between true mean and estimation below some small number) of known distribution
    - 4.5.3.1 how basing the algorithm on different ‘easy’ distribution effects its performance
  - 4.5.4 strength and limitations of each algorithms
- 4.6 overview of inference problems
- 4.7 benchmark inference algorithm by pymc
  - 4.7.1 speculation/analysis of how pymc estimates posterior distribution
- 4.8 how basic sampling algorithms could be expanded to inference algorithms
  - 4.8.1 their performance with respect to the benchmark with different problems
- 4.9 multivariate statistical test
  - 4.9.1 how empircal cdf is calculated for n dimensional distribution
  - 4.9.2 two ways in which ks test was extended to n dimensional case and their limitations (need to ask Fred)
  - 4.9.3 how n dimensional distributions are tested in the code base (with respect to exact cdf/ exact sample)
- 4.10 overview of the second problem
  - 4.10.1 benchmark problems to test the unknown inference algorithm