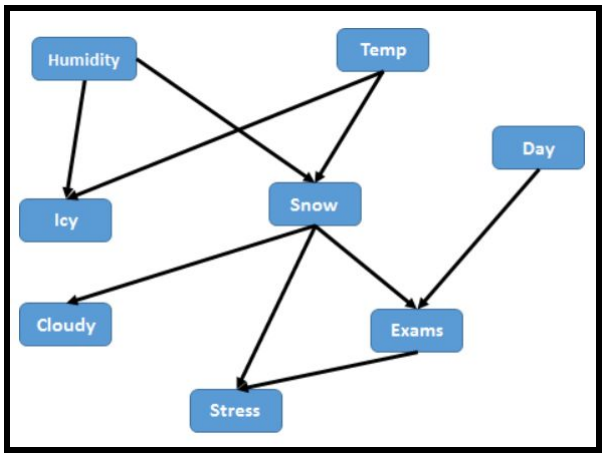
Tucker Haydon
Arthur Lockman
John Lomi
Jon Sawin

AI Project 3

17th February 2017

OVERVIEW

For this assignment, we implemented rejection sampling to estimate probabilities of events occurring in the bayesian network shown below. Each node on the graph has its own set of probabilities, conditional on its parents' states. Rejection sampling generates states for the independent events (humidity, temperature, and day in this case) and then propagates the states through the tree. It then prunes the events that do not match the given conditions (for example, humidity must be high). Finally it averages the number of simulations for a queried node (for example, stress=high).



PART 1: WARMUP

Given the following probabilities:

P(good quiz grades) = 0.8
P(pass) = 0.9
P(good quiz grades | pass) = 0.8

What is P(pass | good quiz grades)?

P(good quiz grades | pass) * P(pass) / P(good quiz grades) = 0.9

What is P(pass ^ good quiz grades)?

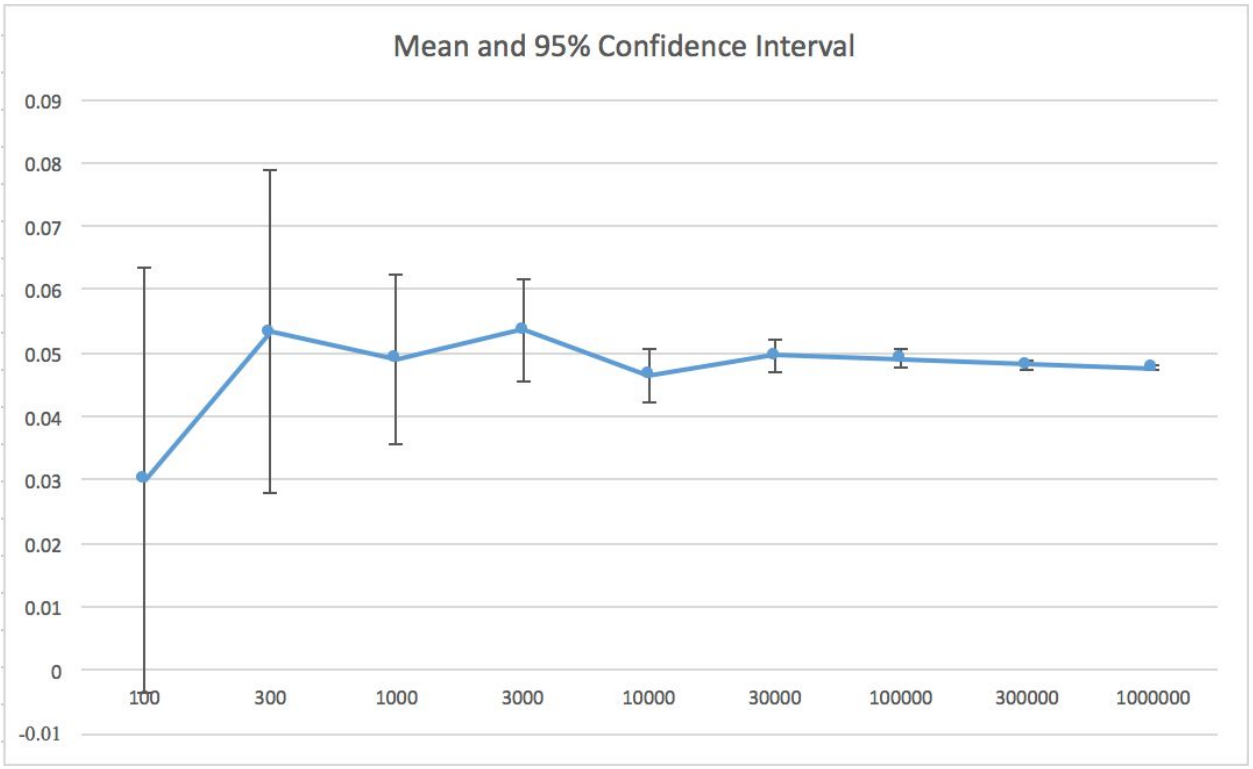
P(pass | good quiz grades) * P(pass) = 0.9 = 0.72

PART 2: REJECTION SAMPLING

We performed rejection sampling using the following iterations: [100 300 1000 3000 10000 30000 100000 300000 1000000] and graphed the mean and 95% confidence interval. The 95% confidence interval for this program was calculated using the Binomial Proportion Confidence Interval: https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval

Test cases:

1. Querying stress=high, no observations



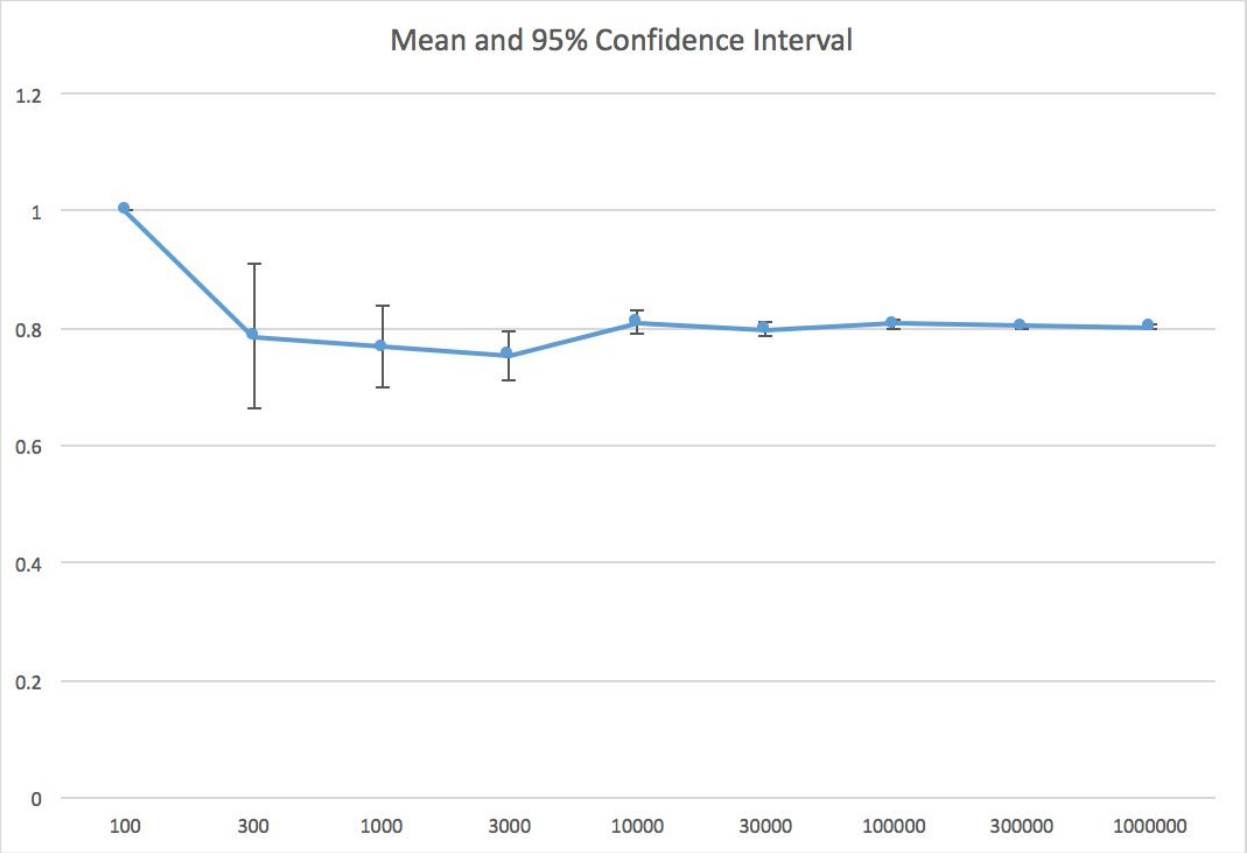
a) The estimates begin to converge between 300 and 1k samples. At this point, the mean value does not change by much more than 0.01 for the rest of the program. It took our machine between 10 and 27 ms to generate enough samples.

b) Confidence Intervals

   i) 0.2: Within the first 100 samples

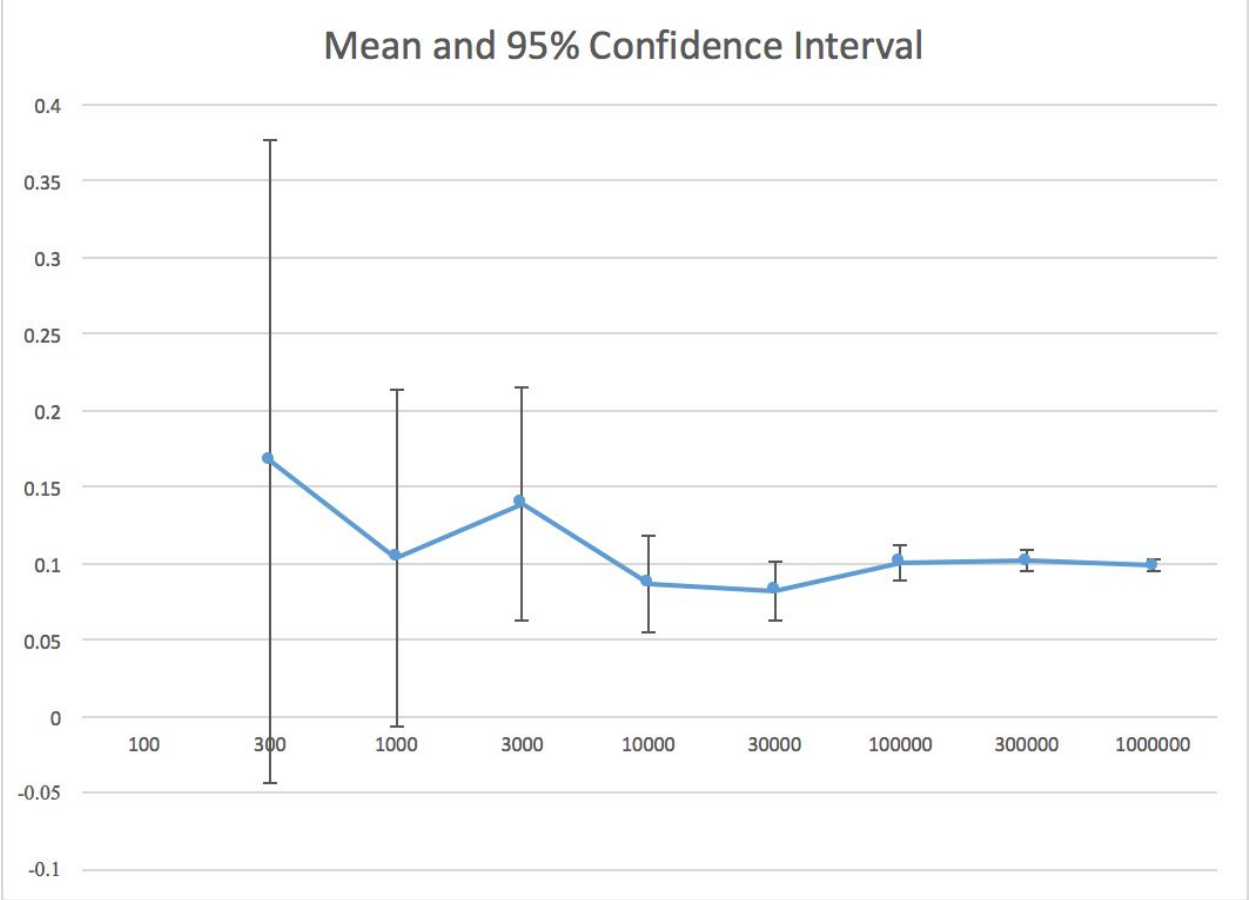
ii) 0.1: Within the first 100 samples
iii) 0.05: Within the first 100 samples
iv) 0.01: Between 1k and 3k samples

2. Querying stress=low, snow=true



a) The estimates begin to converge at around 10k samples. It took our machine around 279 ms to generate enough samples.
b) Confidence Intervals
   i) 0.2: Between 100 and 300 samples
   ii) 0.1: Between 300 and 1000 samples
   iii) 0.05: Between 1k and 3k samples
   iv) 0.01: Between 30k and 100k samples

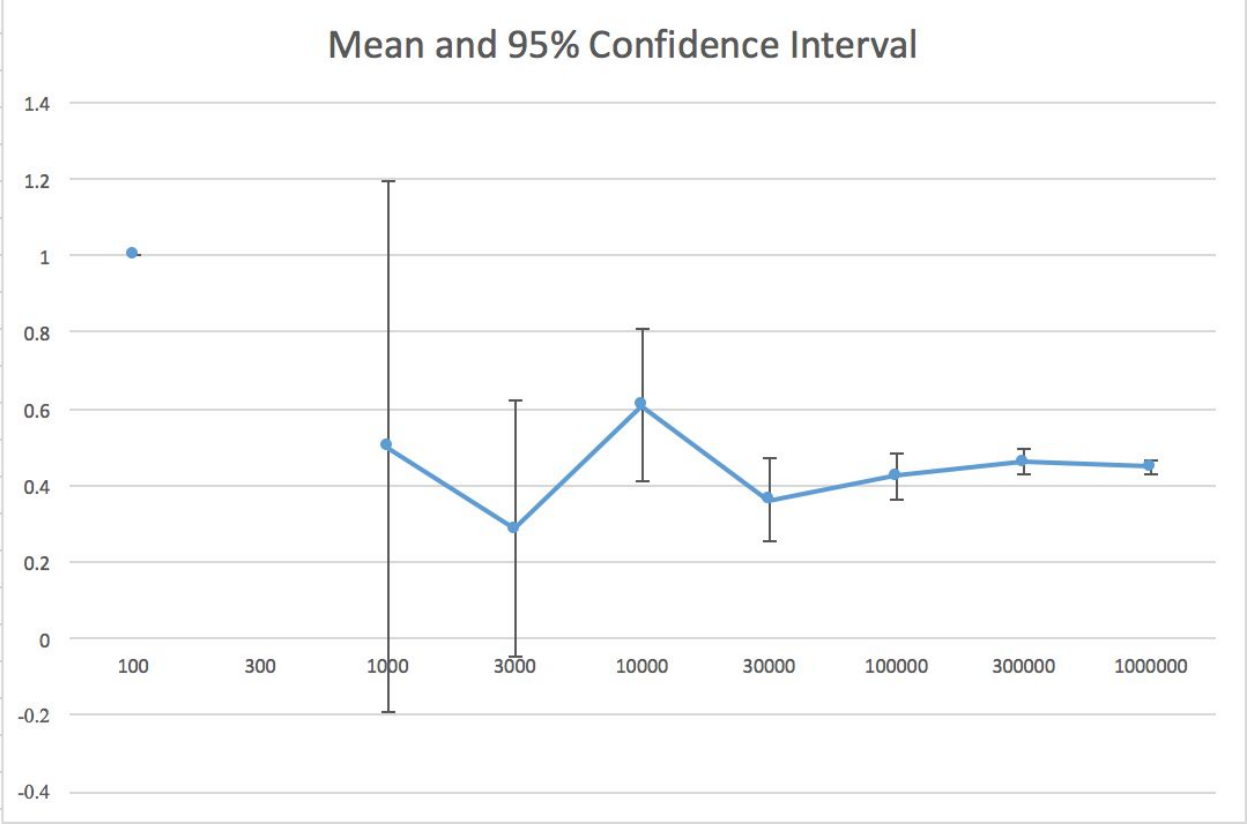
3. Querying stress=high, snow=true, day=weekend



NOTE: For this simulation, no acceptable samples were generated using 100 trial runs

a) The estimates begin to converge around 100k samples. It took our machine around 2734 ms to generate enough samples.
b) Confidence Intervals
    i) 0.2: Around 300 samples
    ii) 0.1: Between 1k and 3k samples
    iii) 0.05: Between 3k and 10k samples
    iv) 0.01: Between 100k and 300k samples

4. Querying humidity=high, snow=true, day=weekend, stress=high



NOTE: For this simulation, no acceptable samples were generated using 300 trial runs

a) The estimates begin to converge around 100k samples. It took our machine around 2806 ms to generate enough samples.
b) Confidence Intervals
    i) 0.2: Around 10k samples
    ii) 0.1: Around 30k samples
    iii) 0.05: Between 100k and 300k samples
    iv) 0.01: Not within the first 1 million samples