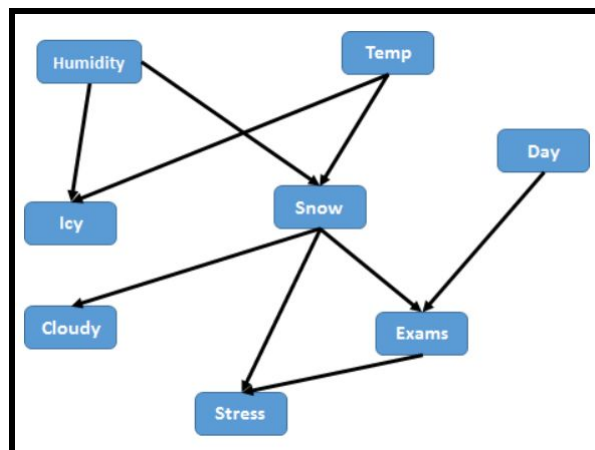

Tucker Haydon
Arthur Lockman
John Lomi
Jon Sawin

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:

$P(\text{good quiz grades}) = 0.8$

$P(\text{pass}) = 0.9$

$P(\text{good quiz grades} \mid \text{pass}) = 0.8$

What is $P(\text{pass} \mid \text{good quiz grades})$?

$$P(\text{good quiz grades} \mid \text{pass}) * P(\text{pass}) / P(\text{good quiz grades}) = 0.9$$

What is $P(\text{pass} \wedge \text{good quiz grades})$?

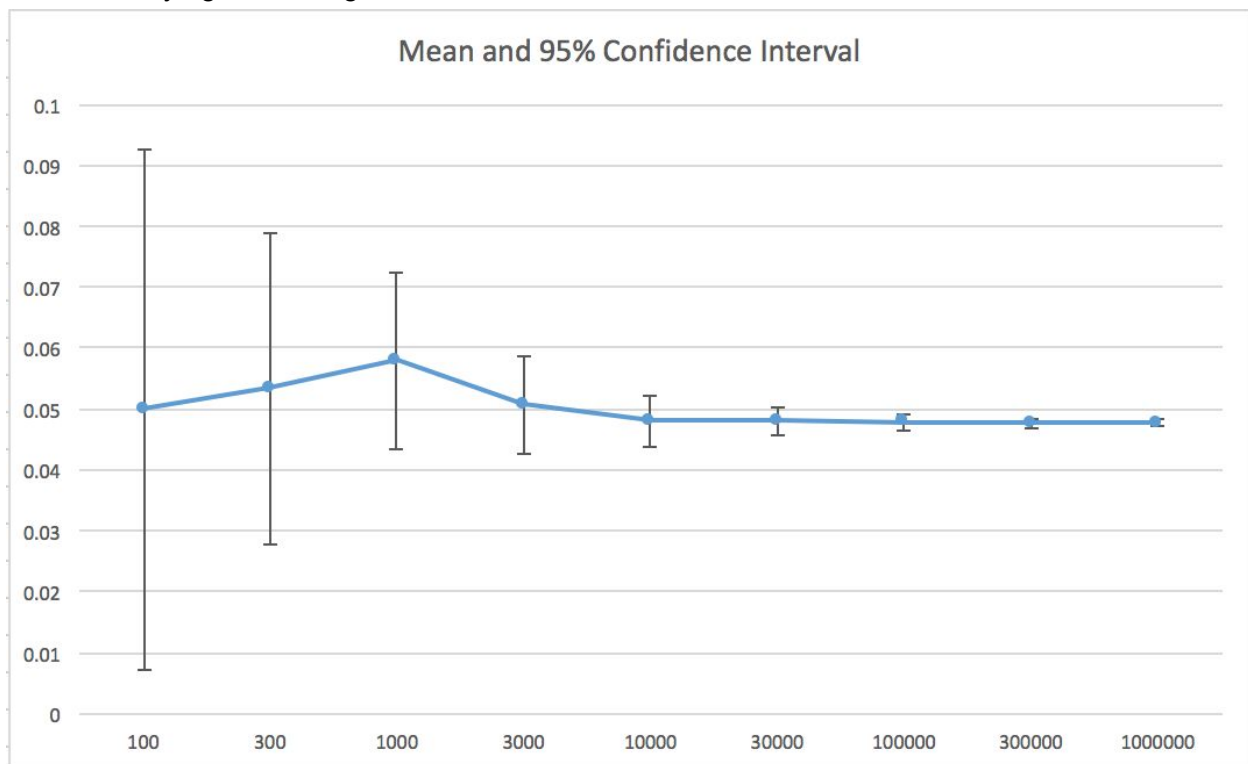
$$P(\text{pass} \mid \text{good quiz grades}) * P(\text{pass}) = 0.9 = 0.72$$

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:

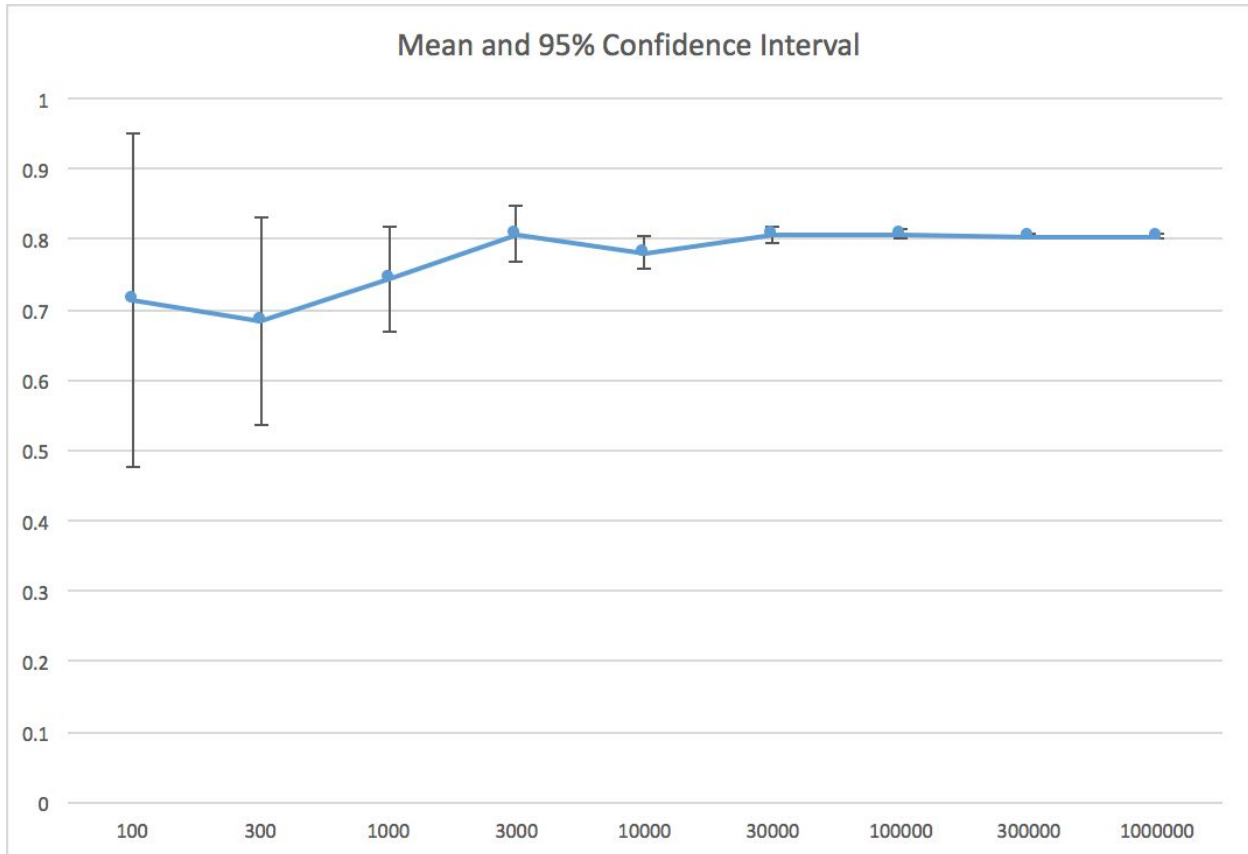
1. Querying stress=high, no observations



- a) The estimates begin to converge between 3k and 10k samples. It took our machine between 91 and 294 ms to generate enough samples.
- b) Confidence Intervals
 - i) 0.2: Within the first 100 samples
 - ii) 0.1: Within the first 100 samples

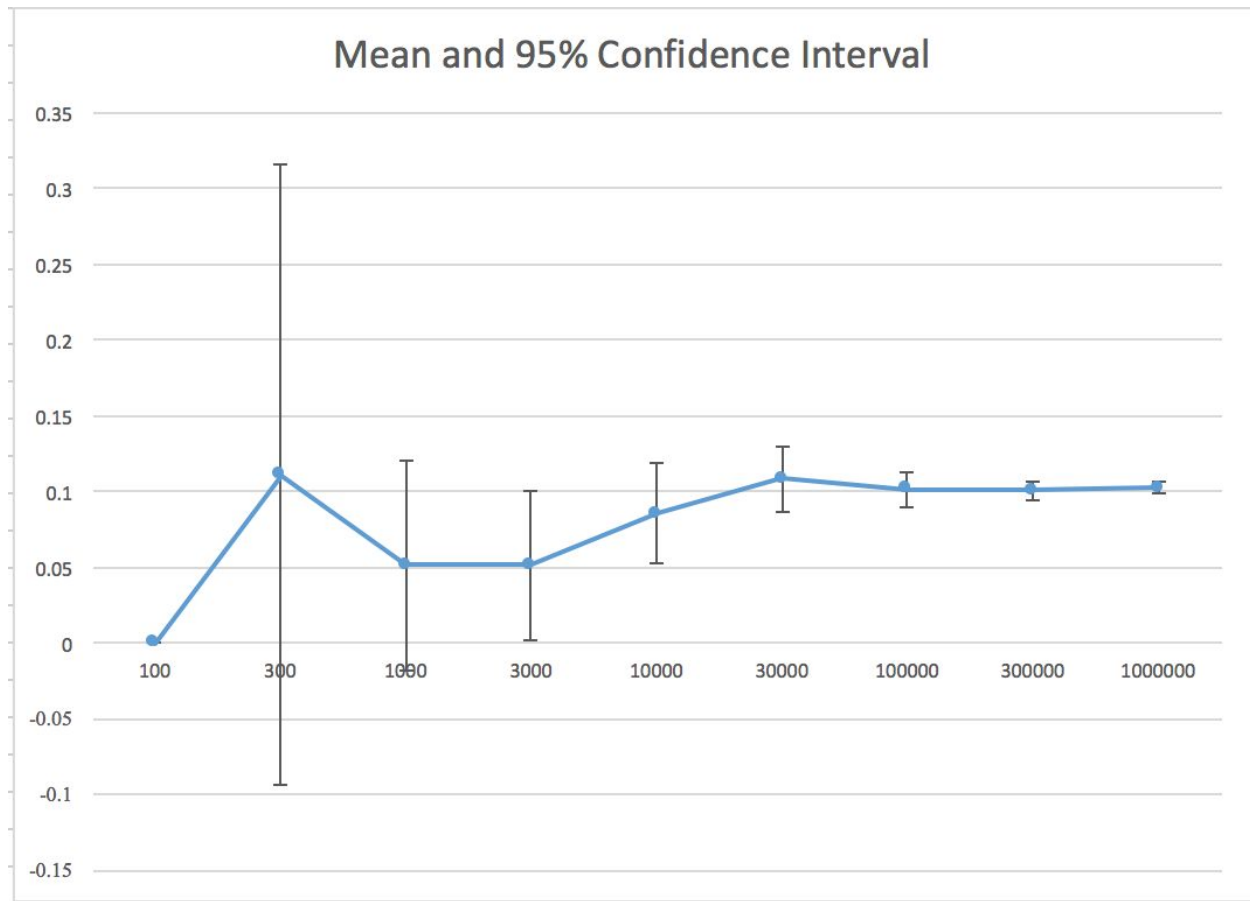
- iii) 0.05: Within the first 100 samples
- iv) 0.01: Between 1k and 3k samples

2. Querying stress=low, snow=true



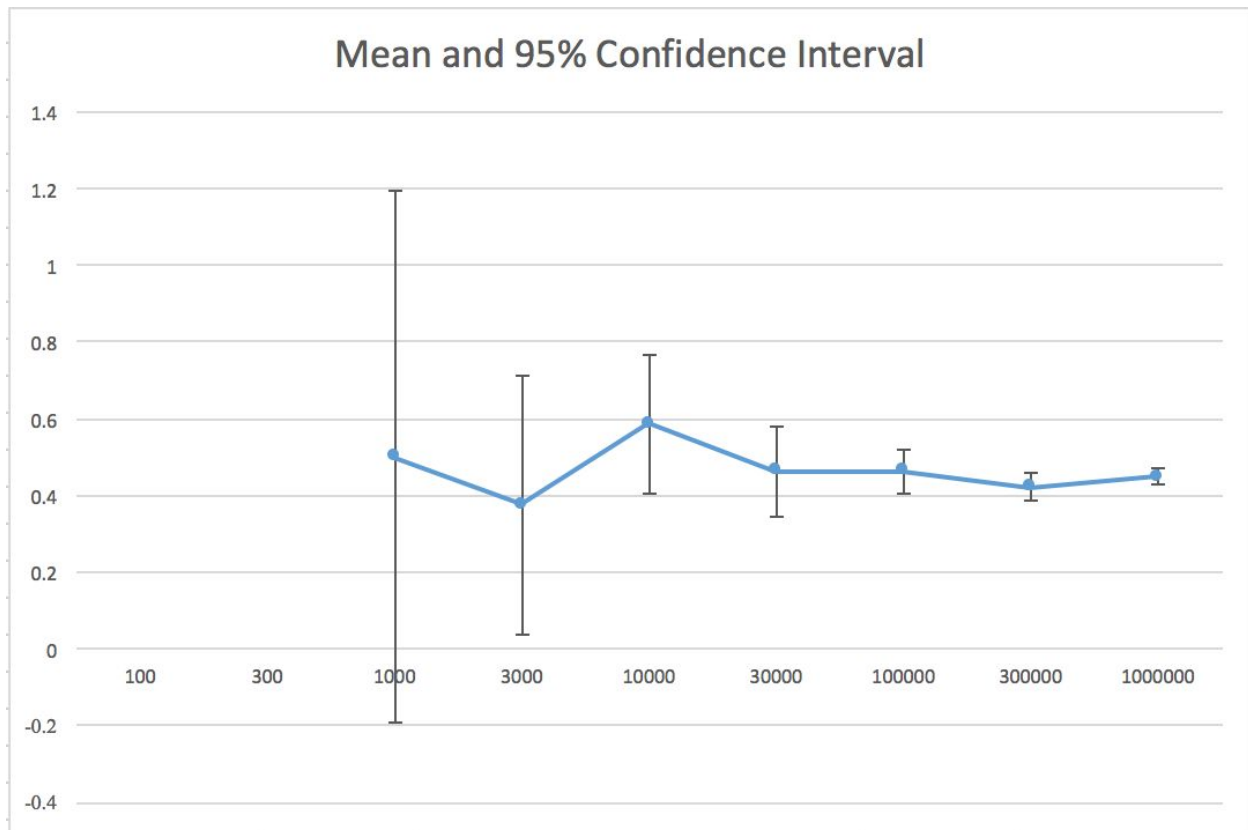
- a) The estimates begin to converge between 3k and 10k samples. It took our machine between 88 and 284 ms to generate enough samples.
- b) Confidence Intervals
 - i) 0.2: Between 100 and 300 samples
 - ii) 0.1: Between 300 and 1000 samples
 - iii) 0.05: Between 1k and 3k samples
 - iv) 0.01: Between 30k and 100k samples

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



- a) The estimates begin to converge between 30k and 100k samples. It took our machine between 828 and 2773 ms to generate enough samples.
- b) Confidence Intervals
 - i) 0.2: Around 300 samples
 - ii) 0.1: Between 300 and 1000 samples
 - iii) 0.05: Between 1k and 3k 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 100 or 300 trial runs

- a) The estimates begin to converge between 30k and 100k samples. It took our machine between 838 and 2852 ms to generate enough samples.
- b) Confidence Intervals
 - i) 0.2: Between 3k and 10k samples
 - ii) 0.1: Between 30k and 100k samples
 - iii) 0.05: Around 100k samples
 - iv) 0.01: Not within the first 1 million samples