Comp9418 – Report – Tasks 2 & 3

Daniel Constantinidis

z5162121

*Task 2*

Since most of this task had already been implemented in the tutorials, there was not much challenge to this task.

The size of the joint distribution with 16 nodes is **2654208*.***

We find this by multiplying the size of the outcome space of each variable by each other. For example outcomeSpace = (A: (0,1), B: (0,1,2), C: (0,1,2,3))

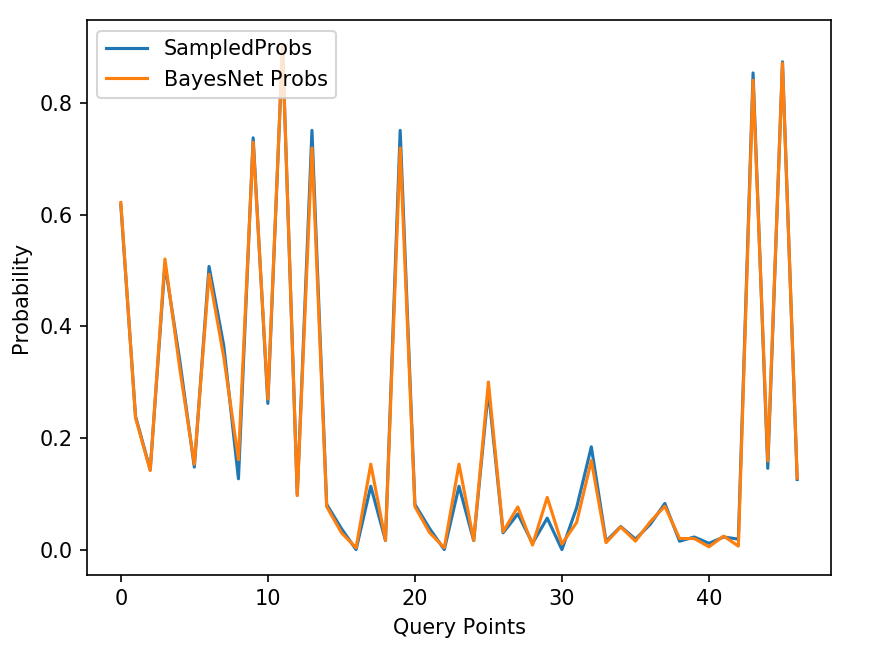
We get 2 \* 3 \* 4 here, multiplying the size of each variables outcome space.

*Task3*

Main Challenges for this task included setting up the data structures for getting the sampled probabilities, as well as the Bayes Net queries. It was tricky, since we needed to collect probabilities for each outcome combination and compare each one. This was resolved by using the structure of the conditional probability tables, and accessing each dictionary item as a probability, with the keys being the outcome. This way, we can access each probability directly given the outcomes and easily compare both probabilities.

The time complexity of the sampling procedure is O(N\*M), N being the number of variables, and M being the number of samples generated. We need to access each node once per sample generation, and we are sampling M times from the graph.

The accuracy of the sampled estimates are pretty close to that of the Bayesian Network for the given queries. We manually created 10 queries, and compared them at each outcome probability.



A Root Mean Squared Error metric was used to determine how closely the sampled estimates represent the ground truth probabilities from the Bayes Net.

\begin{displaymath}RMS Errors= \sqrt{\frac{\sum_{i=1}^n (\hat{y_i}-y_i)^2}{n}}\end{displaymath}

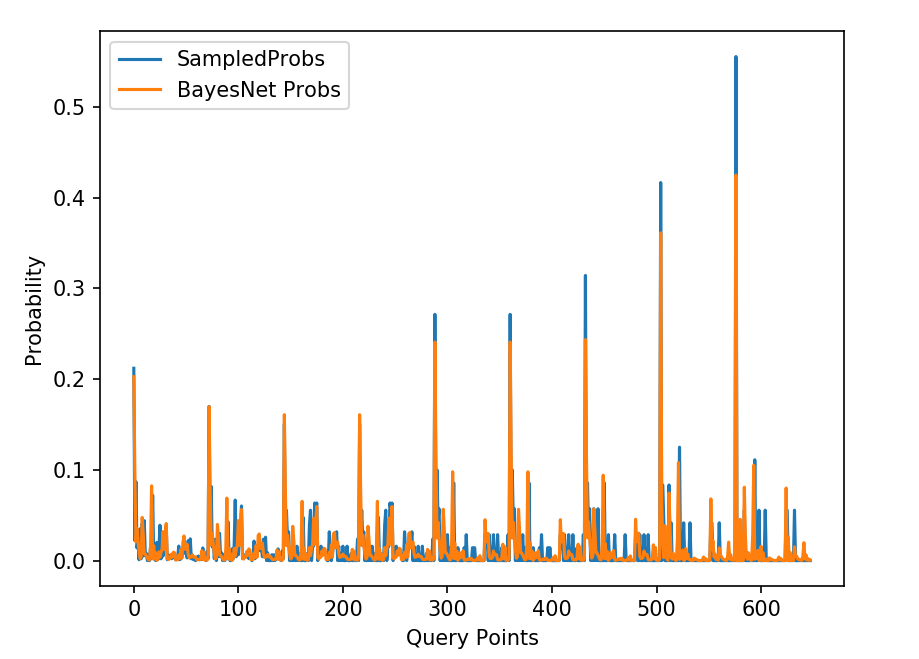
For this experiment the results were;

RMSE = 0.02569876727287478

We used queries with varying amount of ‘evidence’, in order to get a general estimate of the accuracy of the sampling.

As you add more observed variables in the query, the effective sample size reduces. This is because we are discarding samples that disagree with the evidence, and only considering ones where the evidence agrees with the sample. If we add more and more conditions on the variable outcomes, it becomes less likely you will obtain a sample with those observations.

Since we are reducing the effective sample size the accuracy of the sampled estimates is reduced, because there are less probability points which we can use in the comparison, resulting in a potential bias from the reduced sample size. For example, if the probability to get heads was 0.1, and you flipped once, you are very likely not to see a head, as opposed to if you were to flip the coin 100 times. This is evident in the following experiment.



RMSE = 0.023700893359826652

We created 9 queries, with the same variables, but each with an increasing number of observed variables. The further along we go on the x-axis, the more variables we observe. We can see where each new query begins at each ‘spike’ in the graph.

We see that as you increase the evidence, the sampled probabilities disagree more and more with the Bayes Net ground truths.