Compare Likelihood Weighted Sampling to Gibbs Sampling on different Bayesian Network structure and distribution

Experiments set up

- **The criteria of evaluation:**

I will use the Mean of Absolute error of the difference between the queries from original bayesian network (BN) the and the queries from the bayesian network learned from samples.

- **How to generate the CPT of bayesian network:**

Used Dirichlet distribution. There are two version, one is skewed distribution, with geometric sequence of five for alpha series. For example, 10, 5, 1 to give the three dimension (K=3), and I default X_i=0 with alpha 10. Other version is more uniform distribution, set all alpha to 2, and it is symmetric Dirichlet distribution. Also, to let the experiment be more deterministic, I set all the variable with same amount of assignments, set it to 3, that is 0, 1, 2.

Note: Sometime the query will not appear in the samples, I count it with 0. Also, I avoid it with X_i=0 which is the assignment of maximum alpha.

- **There are few cases of bayesian network structure for our experiments**

Case 1:

line shape of bayesian network (each node in degree and out degree is one, except the first node and last node), $X_1 \rightarrow X_2 \ldots\ldots \rightarrow X_{10}$

Case 2:

There are few components in the bayesian network. For simplify, I just used two components, and each is different randomly generated graph. One has 4 edges, another have 9 edges.

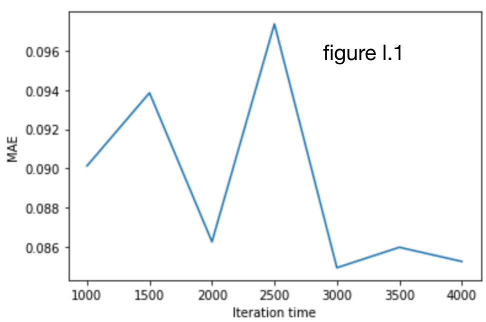
The adjacent link list is:

```
{0: [3], 1: [0], 2: [3, 0], 12: [7, 8], 6: [10], 9: [10], 11: [9, 6, 8], 8:
[10], 13: [9]}                                                *Note i means Xi
```

Likelihood weighted sampling (LWS)

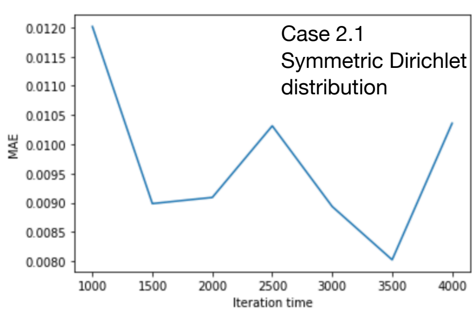
If the variable in the evidence are all root of the bayesian network, then it will give same weight to the each sample, because it dose not have parent. It is the special case for likelihood weight sampling, and it is same as Forward sampling. Yet, it is not a interesting topics to discuss, so I skip this part and set the evidence with non-root variables for likelihood weight sampling. Also, we know the more x_i in the evidence, the weight will be smaller because it multiply the probability more times. It weighted more at the samples which has first observed x_i in the topological sort (root), and less weight to the descendant.

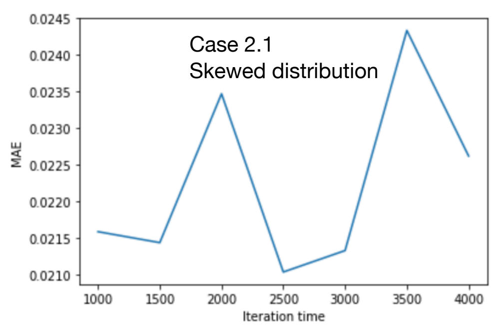
With the case 1, I start from P(X_{10}=0 | X_9) with symmetric Dirichlet distribution, more uniform one. I set the sampling number from 1000 to 6000, increase the 500 each time. When I changed to skewed distribution the starting error got higher, and fluctuated easily. Not very surprise, I also do the P(X_9 | X_{10} =0) on uniform distribution like figure l.1, to know the result of weighted on prior.



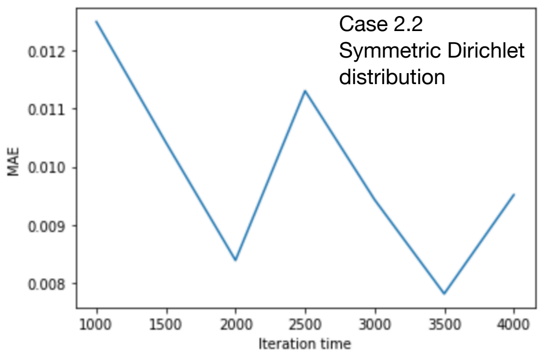
I extended to more evidence, adding X_8, X_7 (structure is $X_7 \rightarrow X_8 \rightarrow X_9 \rightarrow X_{10}$) in it, the MAE became smaller, but the result of dose not improve a lot. For further experiment, I extended to other graph and different combination of evidence and structures.

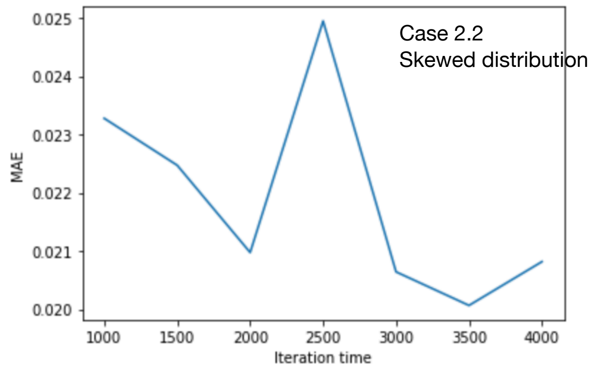
As the case 2, show it below, and I let the scale of plot be closer to Gibbs Sampling. **Case 2.1**. X_9 is the parent of X_{10}, Query P(X_{10}=0 | X_9 , X_{11}), X_{11} is X_9 's parent:



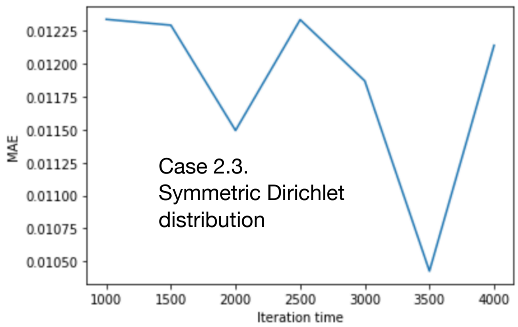


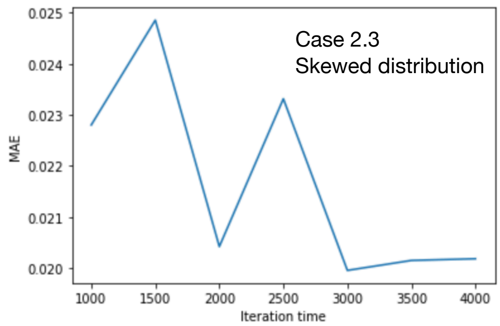
Case 2.2 Query P(X_{10}=0 | X_9 , X_0), X_9, X_0 are in different components:





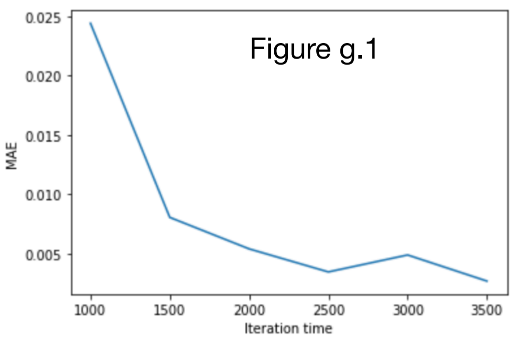
Case 2.3 Query P(X_{10}=0 | X_9 , X_6), X_9, X_6 are both the parents of X_{10},:

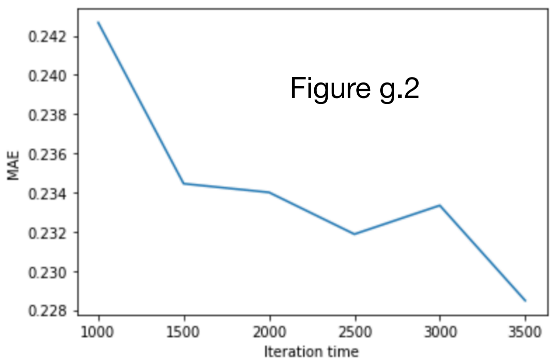




Gibbs sampling

The burn in iteration for the Gibb sampling is actually an hard analysis part, it is not stable and hard to converge, I will not put much on it. The first experiment of gibbs sampling, I start from the query P(X_{10}=0 | X_9) on skewed distribution with case 1's bayesian network structure. The plot of MAE will be Figure g.1.



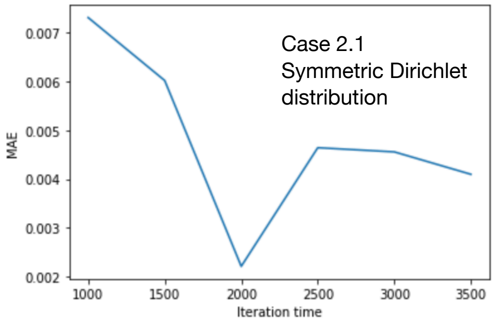


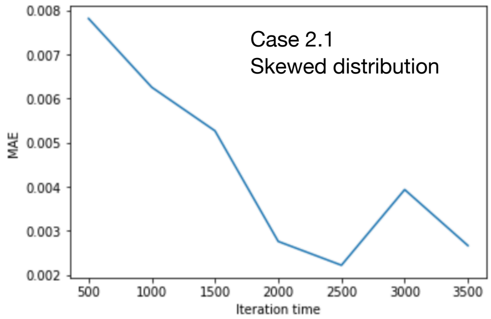
As you can see the line is rise and fall, the error rise and down severely. However, if I change the CPT distribution to the uniform density, the result will not go by our usually expected, it is similar to the result of skewed distribution.

I switch the other query P(X_9=0 | X_{10}) on uniform distribution, in figure g.2, want to make comparison with likelihood weighted sampling, it seems that the performance is not well. However, it keeps decreasing, maybe more iteration it will work out. It can implies that Gibbs sampling's each sample may not have a significant weight on it. I will go further to includes the Case 2 and extend from case 1 to see whether gibbs sampling can out perform in skewed distribution, and make more detail comparison of the different bayesian structure.

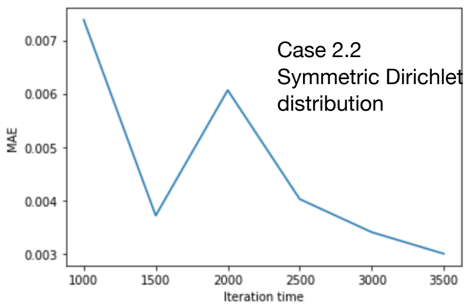
Note:
1.x of plot is the number of burn in iteration, and the number of samples is 50% of it, this is the fixed setting - (G.1)
2. Observe from 500 burn in iteration for skewed distribution

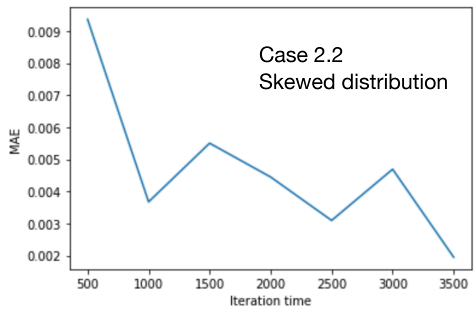
Case 2.1 X_9 is the parent of X_{10}, Query P(X_{10}=0 | X_9 , X_{11}), X_{11} is X_9 's parent:



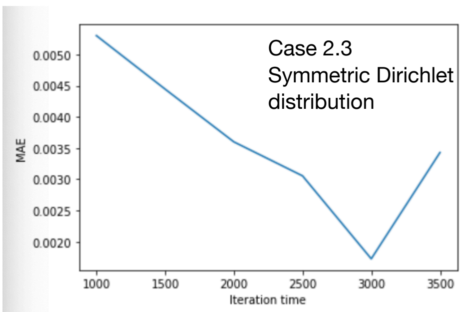


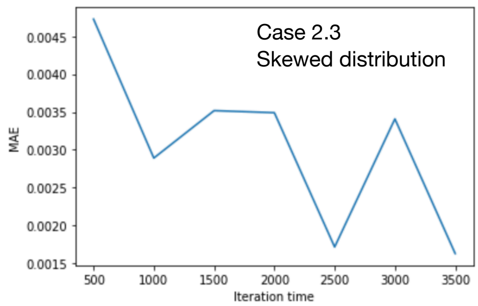
Case 2.2 Query P(X10=0 | X9 , X0), X9, X0 are in different components:





Case 2.3 Query P(X_{10}=0 | X_9 , X_6), X_9, X_6 are both the parents of X_{10}:





Conclusion:

The Likelihood Weighted Sampling **(LWS)** are easier to be effected by the distribution, when it faced skewed distribution, it really easy to fluctuated, and no sign of converge. The range of the MAE is also a lot of difference. However, the Gibbs Sampling **(GS)** can perform converge in the case, and the performance is similar in both distribution. Also, the Gibbs sampling performs better than LWS when one variable in the evidence are independent to the queried variable (by Case 2.2), or variables in the evidence are independent (by Case 2.3). Actually, GS dose not need to sampling by topological order, so it is better in these cases.

On the other side, the LWS has a better performance than GS when evidence on the leaf (by Case 1), by comparing Figure g.2 and Figure l.1. Though LWS can not beat GS in some cases, it has a well performance given parent and parent's parent (by Case 2.1). Also, it has more desirable MAE in Symmetric Dirichlet.

To summarize, GS is not influenced a lot by skewed distribution, even get better result because easily to get closer to the highest probability, and beat LWS in the independent variables case. The LWS is more closer to prior (more weight) , or said depend on prior compared to GS . Thus, it can attain good performance on the non skewed distribution, because I sample all the assignments of evidence. Also, If the queried variable and evidence are dependent, or evidence on the leaf, LWS can get a better performance.