AI project 6
Mengwen Li(mli2)
Yihong Zhou(yzhou8)

Extra Credit:
Part1:
We did the convergence test for both reject sampling and likelihood weighted sampling for each query. The result is as following:

Query1 Reject Sampling Convergence Test

	200	400	600	800	1000
1	0.7238602	0.7238897	0.7232891	0.72418857	0.72374994
2	0.72314787	0.72361726	0.7231997	0.72356457	0.723753
3	0.72375923	0.7235605	0.7237595	0.7243305	0.72345275
4	0.72365206	0.7241065	0.7238643	0.7241022	0.7235797
5	0.72387177	0.7235352	0.7238303	0.7234088	0.7235862
6	0.7234997	0.723681	0.72419524	0.7239835	0.7236238
7	0.7234997	0.7231027	0.7236111	0.724029	0.72367775
8	0.72336924	0.7236771	0.7234702	0.723953	0.72388446
9	0.7242138	0.7237328	0.72344685	0.72355896	0.7236698
10	0.7233063	0.7243572	0.72362596	0.7240092	0.72357625
MEAN	0.723617987	0.723725996	0.723629225	0.72391283	0.723655365
VARIANCE	0.000317489	0.000339702	0.000296667	0.000301162	0.000120411

Query1 Likelihood Weighted Convergence Test

	200	400	600	800	1000
1	0.7384896	0.73200804	0.73497677	0.73420167	0.737095

2	0.72982514	0.73184174	0.7355458	0.7317123	0.73269
3	0.72503805	0.7354632	0.74006957	0.73196584	0.73540044
4	0.7349918	0.7352832	0.73503226	0.73638254	0.73263085
5	0.73426056	0.7326278	0.7321926	0.7352507	0.73427093
6	0.73504233	0.73666906	0.7339623	0.7325563	0.7364426
7	0.73504233	0.7331916	0.73374903	0.7356243	0.73680323
8	0.735747	0.7404542	0.7370691	0.7371723	0.73561466
9	0.73696154	0.73393357	0.7327135	0.736384	0.73495007
10	0.7429273	0.73177165	0.7378145	0.73586255	0.73560923
MEAN	0.734832565	0.734324406	0.735312543	0.73471125	0.735150701
VARIANCE	0.00478092	0.002741616	0.002426179	0.001986977	0.001559451

Qurey2 Reject Sampling Convergence Test

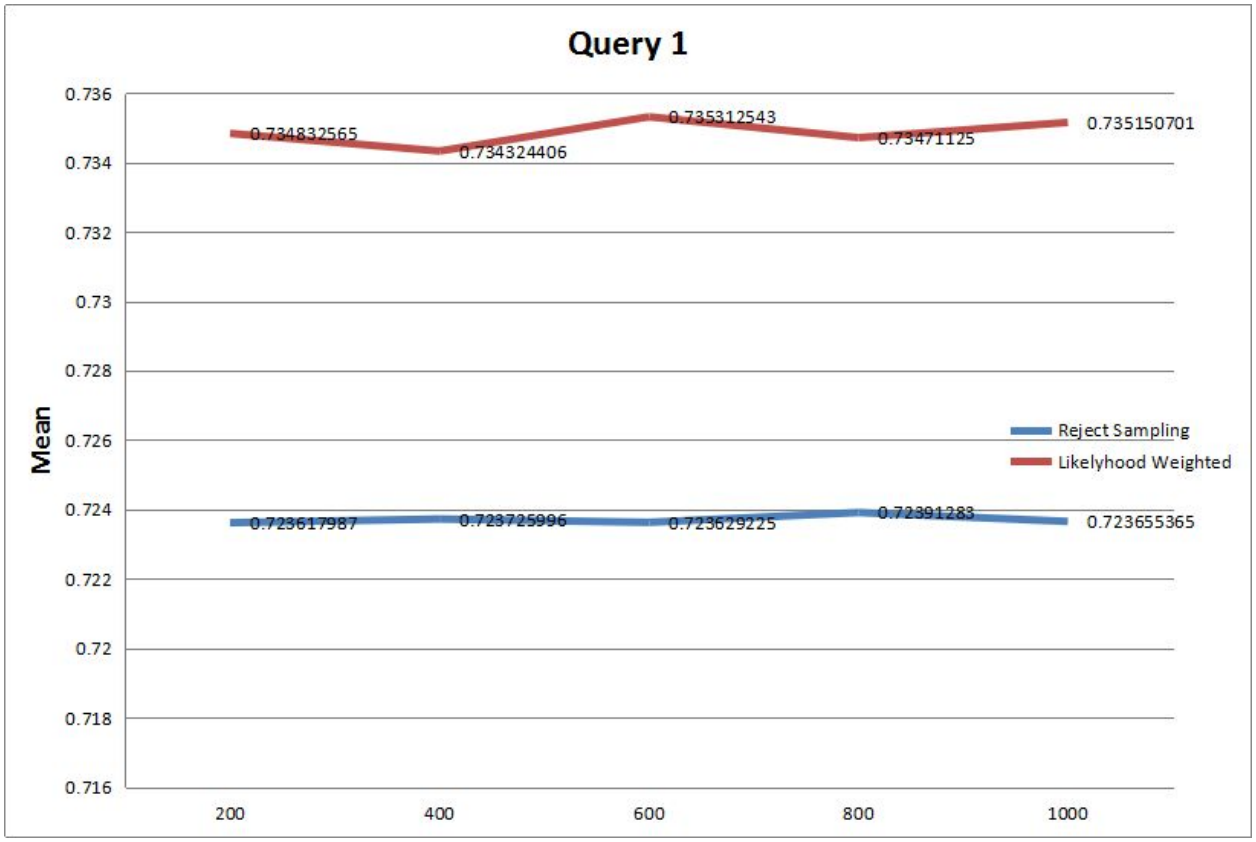
	200	400	600	800	1000
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN

MEAN	0	0	0	0	0
VARIANCE	0	0	0	0	0

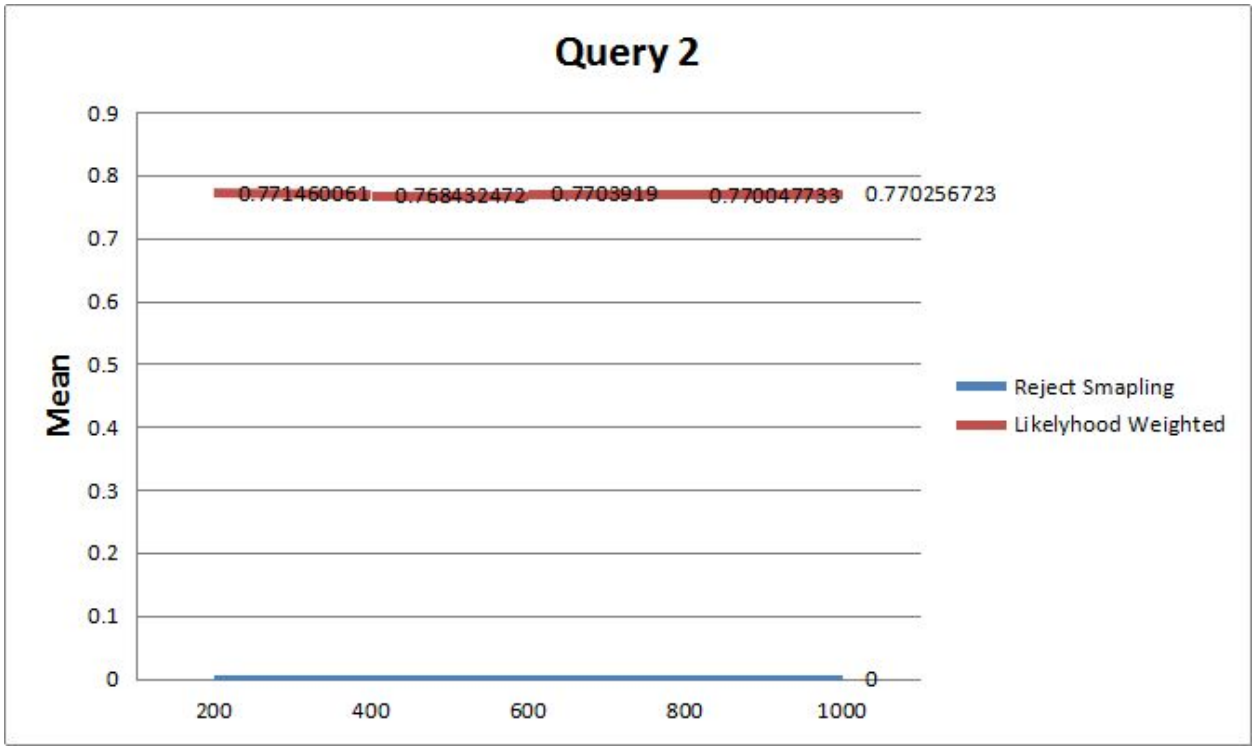
Query2 Likelihood Weighted Convergence Test

	200	400	600	800	1000
1	0.7644	0.765825	0.77168846	0.7699949	0.77192515
2	0.77430016	0.7665749	0.7700554	0.7691328	0.77045697
3	0.7745001	0.76792526	0.7709666	0.7721643	0.76952994
4	0.7662999	0.7699999	0.7717996	0.77113336	0.76997393
5	0.76850015	0.7679503	0.77128804	0.7682823	0.7709168
6	0.7729	0.7722497	0.77168846	0.769464	0.7709737
7	0.7746001	0.76347506	0.76998824	0.7700512	0.7697583
8	0.7760002	0.7695248	0.7695221	0.7713702	0.7690416
9	0.7674	0.7731247	0.766889	0.77130747	0.77027744
10	0.7757	0.7676751	0.7700331	0.7675768	0.7697134
MEAN	0.771460061	0.768432472	0.7703919	0.770047733	0.770256723
VARIANCE	0.004340355	0.00290836	0.001487312	0.001468076	0.000843424

The rejection sampling has allowed variance range 0-0.001 whereas likelihood weighted sampling has variance range 0-0.0073. The variance range of likelihood weighted sampling can be smaller but we didn't try further because it takes super long time to wait for likelihood weighted sampling to converge and reduce its variance range under 0.0073. We believe likelihood weighted sampling can do better to converge down to a more precise probability.



The mean probability of both rejecting sampling and likelihood weighted sampling is shown above. Both algorithm are stable, though rejecting sampling has better performance because its variance is not as big as likelihood weighted sampling. Both result is very similar, only about 0.01% difference among two algorithm result. Both algorithm converges to a probability round 0.73



When dealing with query2, reject sampling probability is always NaN. So the model using rejection sampling is not applicable. The mean of likelihood weighted sampling is very stable. It is always converge to around 0.77.

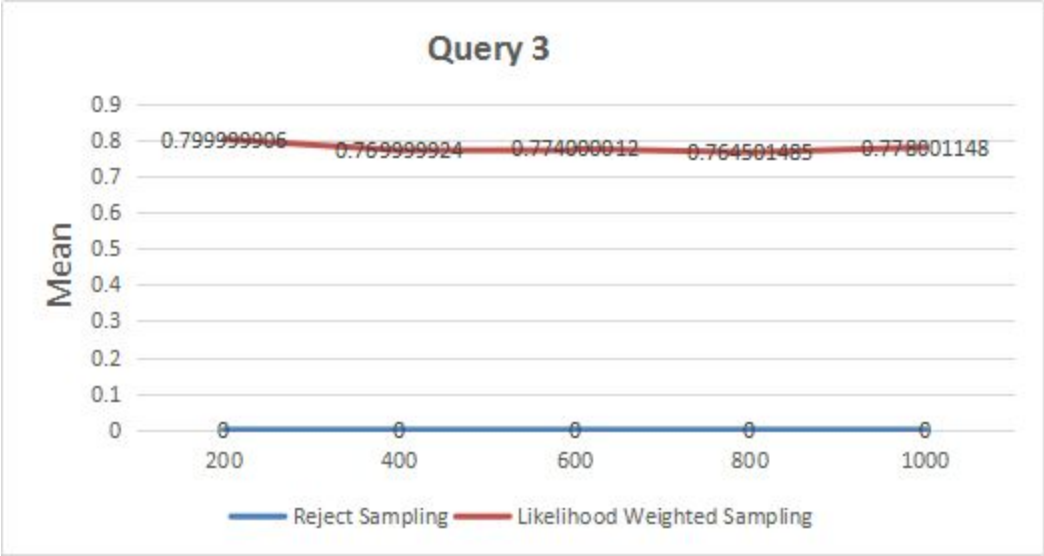
There are other query that leads to likelihood weighting significantly faster.
Query f,f,f,f,f,q,f,f fulfill the requirement.
The result of this query shown as following:
Reject Sampling

	200	400	600	800	1000
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	0
4	NaN	NaN	NaN	0	NaN
5	NaN	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN	NaN

7	NaN	NaN	NaN	0.5	0
8	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	0
10	NaN	0	0	NaN	NaN
MEAN	0	0	0	0.25	0
VARIANCE	0	0	0	0.353553391	0

Likelihood weighted Sampling

	200	400	600	800	1000
1	0.7599999	0.7799999	0.79333335	0.77500147	0.78000116
2	0.81999993	0.7699999	0.7866667	0.7700015	0.808001
3	0.81999993	0.74999994	0.71999997	0.76000154	0.79200107
4	0.7599999	0.7699999	0.76	0.8000013	0.76400125
5	0.8399999	0.7799999	0.76666665	0.7350016	0.77600116
6	0.8399999	0.7699999	0.8333334	0.7400016	0.7280014
7	0.7999999	0.75999993	0.74666667	0.7200016	0.7720012
8	7599999	0.7699999	0.76666665	0.7700015	0.800001
9	0.7599999	0.71999997	0.7866667	0.76000154	0.76000124
10	0.8399999	0.83	0.78000003	0.8150012	0.800001
MEAN	760000.624	0.769999924	0.774000012	0.764501485	0.778001148
VARIANCE	2403330.451	0.027487376	0.030217346	0.028814415	0.023869888



The reject Sampling is not applicable in this situation. The query's probability is so small that even 1000 samples can't have a valid query with valid evidences. So in this case, likelihood weighting significantly faster than reject sampling who remains at 0 and does not converge at all.