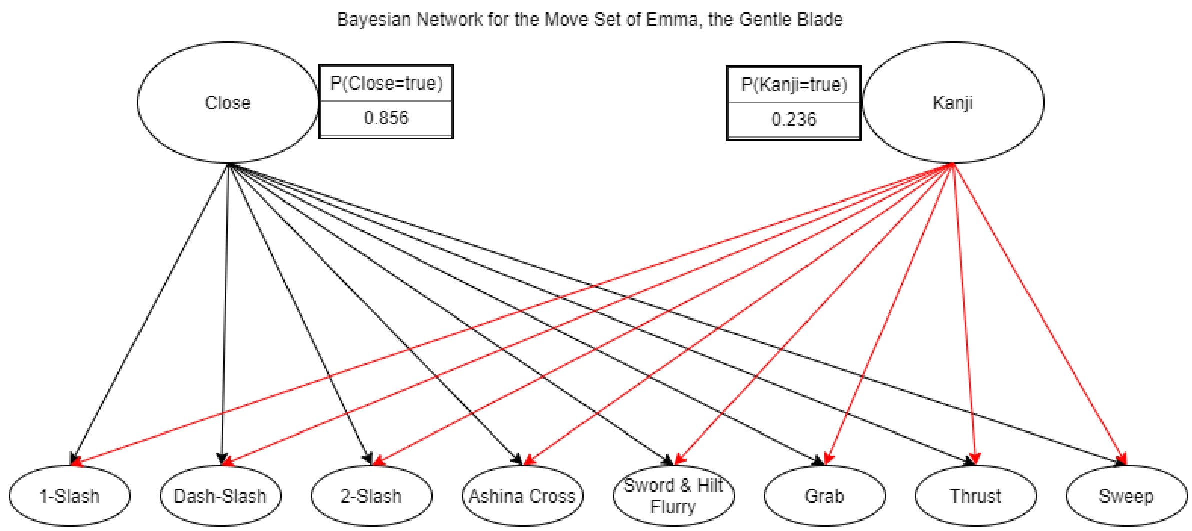
CSC 442: Project 3 - Probability - Brooke Brehm & Ky Potter

Custom Network and Variables



1-Slash			Dash-Slash			2-Slash			Ashina Cross			Sword & Hilt Flurry			Grab			Thrust			Sweep		
C	K	P(1-S=true|C,K)	C	K	P(D-S=true|C,K)	C	K	P(2-S=true|C,K)	C	K	P(AC=true|C,K)	C	K	P(SH=true|C,K)	C	K	P(G=true|C,K)	C	K	P(T=true|C,K)	C	K	P(S=true|C,K)
t	t	0	t	t	0	t	t	0	t	t	0	t	t	0	t	t	0.397	t	t	0.34	t	t	0.263
t	f	0.444	t	f	0.237	t	f	0.159	t	f	0.05	t	f	0.11	t	f	0	t	f	0	t	f	0
f	t	0	f	t	0	f	t	0	f	t	0	f	t	0	f	t	0	f	t	1	f	t	0
f	f	0.744	f	f	0.012	f	f	0	f	f	0.244	f	f	0	f	f	0	f	f	0	f	f	0

(Figure 1. Bayesian Network for the move set of Emma, the Gentle Blade, a boss from the video game Sekiro: Shadows Die Twice.)

The custom Bayesian Network created for this project is based on the move set and determining factors for Emma, the Gentle Blade, a boss from the video game Sekiro: Shadows Die Twice. The moves included are all offensive moves made toward the player in her boss arena. All moves are dependent on the distance (whether the player is close or not - _Close) of Emma from the player and whether a red "warning" kanji (Kanji) symbol appears before the move.

Emma has eight total possible moves. The One_Slash move involves Emma attempting to strike the player with a single slash of her sword, if she is far away the move is preceded by a quick walk toward the player. The Dash_Slash happens when Emma quickly moves past the player from the left or right side while attempting to strike with her sword. Emma's Two_Slash move is marked by Emma attempting two quick slashes of her sword and immediately taking a step back. When Emma sheathes her sword and quickly moves to make a large cross symbol attack, that is the Ashina Cross (Ashina). If she initiates the Ashina Cross while in range for the other close moves, we count it as _Close=True. Her Sword and Hilt Flurry (Sword_Hilt) is a combination of a windup with her sword and four hit attempts. Emma also attempts to grab and flip the player (Grab), thrust her sword into the player (Thrust), and low sweep her sword (Sweep). Emma jumps through the air toward the player during her thrust attack if she initiates the move when she is not close. A video of a fight displaying the move set created for this project can be found here, warning: the game is rated PEGI 18 for violence.

All of Emma's moves can be initiated while close to the player (_Close=T,C=T). However, Emma's Grab, Thrust, and Sweep moves only occur after a red kanji symbol appears on the player's screen (Kanji=T,K=T). Only the One_Slash, Dash_Slash, Ashina Cross, and Thrust moves can be initiated while Emma is not close to the player. Figure 1 above shows the entire network of nodes and probabilities calculated from our collected fight data. The parents of all eight attack moves are Kanji and _Close, so all attack move probabilities depend on them. Kanji and _Close do not have parents, and none of the attack moves have their own children.

In our program, the Bayesian Network class (BNet) is built using a node class (BNode). Each variable (eight moves and two determining factors), is initialized as a node with a name (string), children (list if applicable), parents list, and probability (a dictionary of True/False values and probabilities), with a built-in probability calculating function. The variable nodes are added to a Bayesian Network to initialize it with a network name, list of nodes, and a list of string node names.

Custom Queries

Causal

We tested several causal queries. These include: P(One_Slash|Close=T, Kanji=T), P(Dash_Slash|Close=T, Kanji=T), P(Two_Slash|Close=T, Kanji=T), P(Ashina,|Close=T, Kanji=T), P(Sword_Hilt|Close=T, Kanji=T), P(Grab|Close=T, Kanji=T), P(Thrust|Close=T, Kanji=T), P(Sweep|Close=T, Kanji=T), P(One_Slash|Close=T), P(Dash_Slash|Close=T), P(Two_Slash|Close=T), P(Ashina|Close=T), P(Sword_Hilt|Close=T), P(Grab|Close=T), P(Thrust|Close=T), P(Sweep|Close=T), and P(Kanji|Close=T) . You can see the results of the sampling as well as the enumeration_ask probabilities in Table 1 and Table 2 below.

Table 1: Gibbs Causal Sampling

Enumeration Ask Prob:	0	0	0	0	0	0.397	0.34	0.263
N	One_Slash|C,K=T	Dash_Slash|C,K=T	Two_Slash|C,K=T	Ashina|C,K=T	Sword_Hilt|C,K=T	Grab|C,K=T	Thrust|C,K=T	Sweep|C,K=T
10	0	0	0	0	0	0.6	0.4	0.1
100	0	0	0	0	0	0.42	0.47	0.34
1000	0	0	0	0	0	0.397	0.326	0.293
5000	0	0	0	0	0	0.407	0.3458	0.259
7500	0	0	0	0	0	0.3977	0.3377	0.2577
10000	0	0	0	0	0	0.4003	0.3344	0.2617
50000	0	0	0	0	0	0.3947	0.3398	0.2635
100000	0	0	0	0	0	0.3950	0.3397	0.2628

Enumeration Ask Prob:	0.339216	0.181068	0.121476	0.0382	0.08404	0.093692	0.08024	0.062068
N	One_Slash|C=T	Dash_Slash|C=T	Two_Slash|C=T	Ashina|C=T	Sword_Hilt|C=T	Grab|C=T	Thrust|C=T	Sweep|C=T
10	0.3	0.3	0.3	0	0.1	0	0	0
100	0.39	0.17	0.17	0.02	0.1	0.07	0.04	0.03
1000	0.342	0.192	0.132	0.038	0.075	0.083	0.072	0.061
5000	0.3458	0.194	0.1266	0.0358	0.0832	0.094	0.0824	0.0642
7500	0.3337	0.1863	0.1203	0.0381	0.0844	0.0996	0.0779	0.0613
10000	0.3381	0.1883	0.1267	0.0405	0.0829	0.0938	0.0775	0.0596
50000	0.3363	0.1775	0.1222	0.0374	0.0875	0.0935	0.0798	0.0640
100000	0.3381	0.1810	0.1196	0.0380	0.0836	0.0939	0.0799	0.0631

Table 2: Rejection Causal Sampling

Enumeration Ask Prob:	0	0	0	0	0	0.397	0.34	0.263
N	One_Slash|C,K=T	Dash_Slash|C,K=T	Two_Slash|C,K=T	Ashina|C,K=T	Sword_Hilt|C,K=T	Grab|C,K=T	Thrust|C,K=T	Sweep|C,K=T
10	0	0	0	0	0	0.6667	0.6667	0.0000
100	0	0	0	0	0	0.3684	0.3182	0.2273
1000	0	0	0	0	0	0.3911	0.3726	0.2330
5000	0	0	0	0	0	0.4000	0.3294	0.2574
7500	0	0	0	0	0	0.3978	0.3294	0.2512
10000	0	0	0	0	0	0.3973	0.3689	0.2617
50000	0	0	0	0	0	0.3923	0.3475	0.2576
100000	0	0	0	0	0	0.3977	0.3367	0.2617

Enumeration Ask Prob:	0.3379	0.1804	0.1208	0.0382	0.0841	0.0948	0.0810	0.0627
N	One_Slash|C=T	Dash_Slash|C=T	Two_Slash|C=T	Ashina|C=T	Sword_Hilt|C=T	Grab|C=T	Thrust|C=T	Sweep|C=T
10	0.25	0.2222	0	0.1	0.2222	0.125	0.25	0
100	0.3797	0.1685	0.1264	0.0345	0.1047	0.0814	0.0854	0.0698
1000	0.3520	0.1825	0.1027	0.0299	0.0749	0.0979	0.0761	0.0579
5000	0.3451	0.1875	0.1222	0.0337	0.0852	0.0914	0.0774	0.0654
7500	0.3477	0.1816	0.1189	0.0364	0.0880	0.0905	0.0796	0.0605
10000	0.3390	0.1799	0.1209	0.0366	0.0834	0.0935	0.0812	0.0596
50000	0.3385	0.1823	0.1210	0.0367	0.0852	0.0959	0.0786	0.0619
100000	0.3368	0.1823	0.1201	0.0376	0.0856	0.0938	0.0799	0.0639

Diagnostic

We tested several diagnostic queries. These include: P(Kanji|One_Slash=T), P(_Close|One_Slash=T), P(Kanji|Thrust=T), P(_Close|Thrust=T). You can see the results of the sampling as well as the enumeration_ask probabilities in Table 3 and Table 4 below.

Table 3: Gibbs Diagnostic Sampling

Enumeration Ask Prob:	0	0.7801	1	0.6690
N	Kanji|One_Slash=T	_Close|One_Slash=T	Kanji|Thrust=T	_Close|Thrust=T
10	0	0.7	0.9	0.7
100	0	0.77	0.99	0.69
1000	0	0.784	0.999	0.654
5000	0	0.7914	0.9996	0.6674
7500	0	0.7687	0.9989	0.6668
10000	0	0.7784	0.9984	0.6751
50000	0	0.7769	0.9999	0.6637
100000	0	0.7790	1.0000	0.6672

Table 4: Rejection Diagnostic Sampling

Enumeration Ask Prob:	**0**	**0.7801**	**1**	**0.6690**
N	**Kanji|One_Slash=T**	**_Close|One_Slash=T**	**Kanji|Thrust=T**	**_Close|Thrust=T**
10	0	0.3333	1	0
100	0	0.7333	1	0.6154
1000	0	0.7594	1	0.6944
5000	0	0.7908	1	0.6557
7500	0	0.7711	1	0.6664
10000	0	0.7768	1	0.6684
50000	0	0.7785	1	0.6647
100000	0	0.7831	1	0.6648

Sanity Check

We tested two sanity check queries. These include: P(_Close|Kanji=T), P(Kanji|_Close=T). You can see the results of the sampling as well as the enumeration_ask probabilities in Table 5 and Table 6 below.

Table 5: Gibbs Sanity Check Sampling

Enumeration Ask Prob:	**0.856**	**0.236**
N	**_Close|K=T**	**Kanji|C=T**
10	0.7	0
100	0.9	0.06
1000	0.869	0.31
5000	0.8602	0.2496
7500	0.8608	0.2337
10000	0.8525	0.2346
50000	0.8516	0.2281
100000	0.8572	0.2350

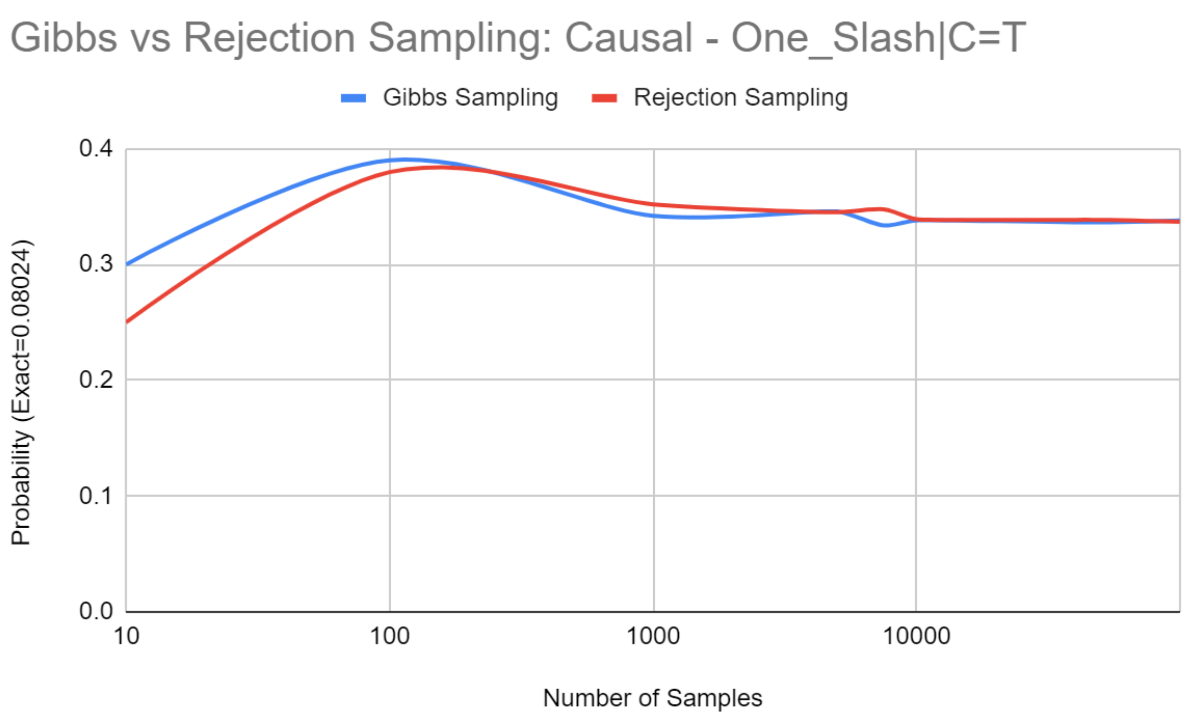
Table 6: Rejection Sanity Check Sampling

Enumeration Ask Prob:	**0.856**	**0.236**
N	**_Close|K=T**	**Kanji|C=T**
10	1	0.3
100	0.95	0.1707
1000	0.8577	0.2151
5000	0.8586	0.2335
7500	0.8576	0.2420
10000	0.8631	0.2302
50000	0.8573	0.2334
100000	0.8519	0.2341

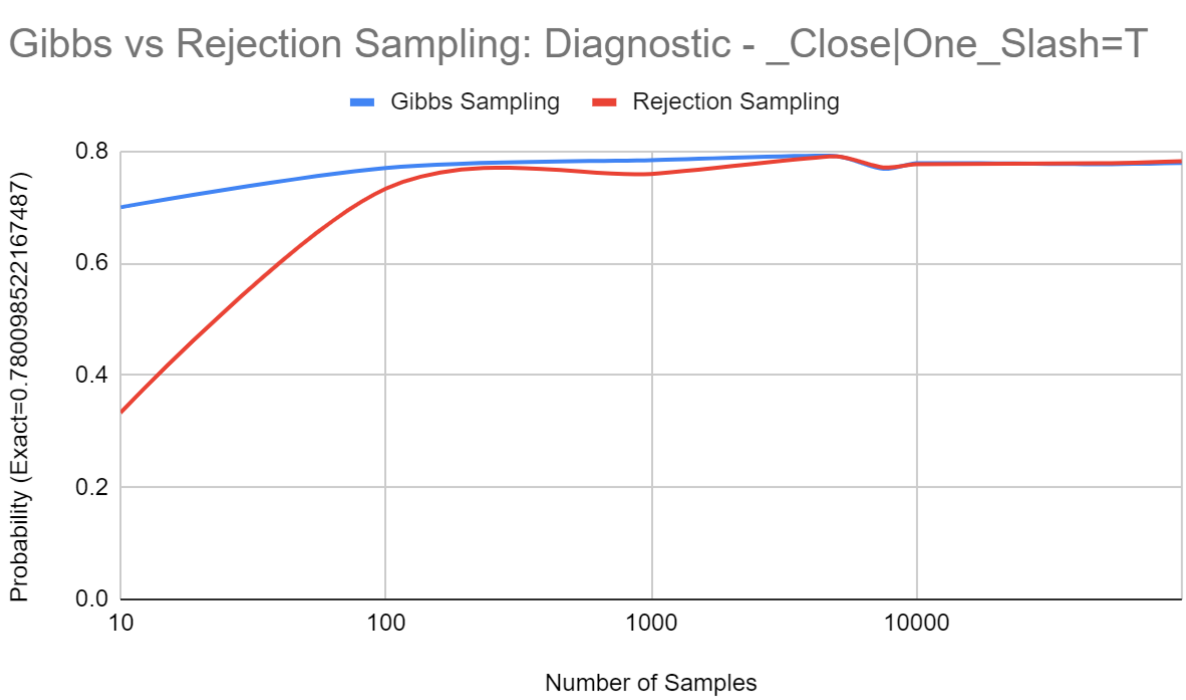
Convergence Graphs

Below are the three convergence graphs comparing the Gibbs sampling and rejection Sampling algorithms we implemented. In Graph 1, the causal query, and Graph 2, the diagnostic query, both Gibbs and rejection sampling probability converge toward the exact probability at 10,000 samples. The sanity check, Graph 3, has Gibbs and rejection sampling probability converging a lot earlier, starting at 1,000 samples.

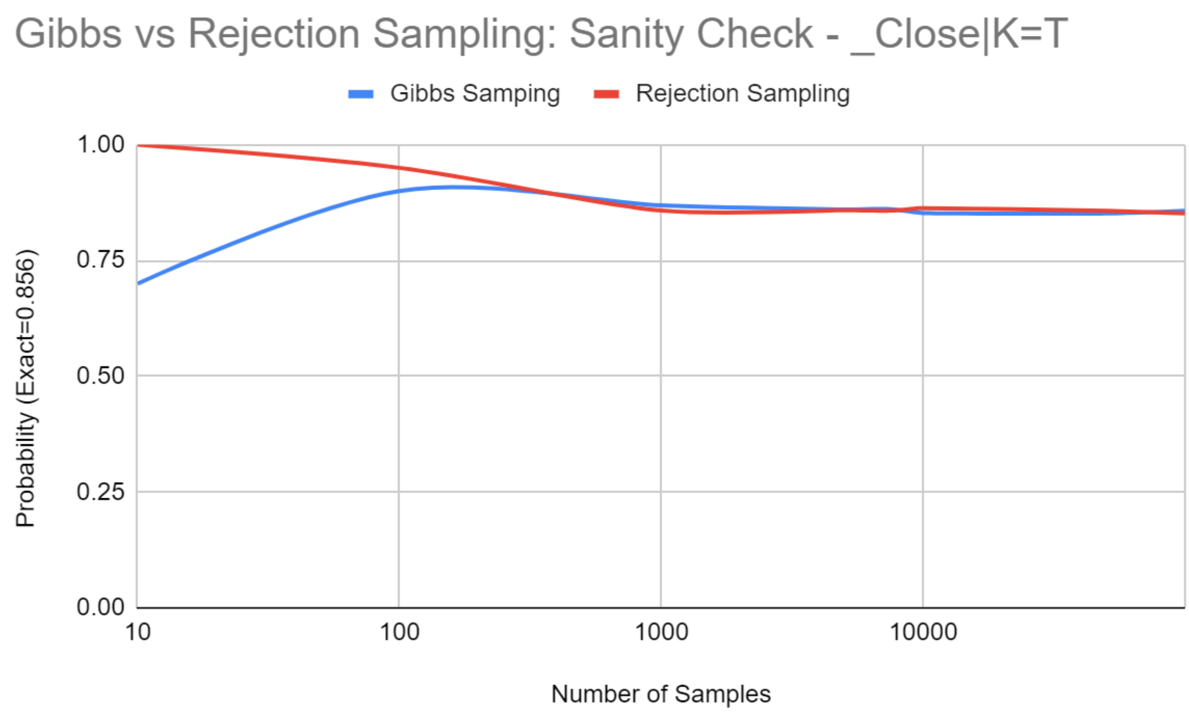
Graph 1: Gibbs and Rejection Sampling Causal Query Convergence for One_Slash|C=T



Graph 2: Gibbs and Rejection Sampling Diagnostic Query Convergence for _Close|One_Slash=T



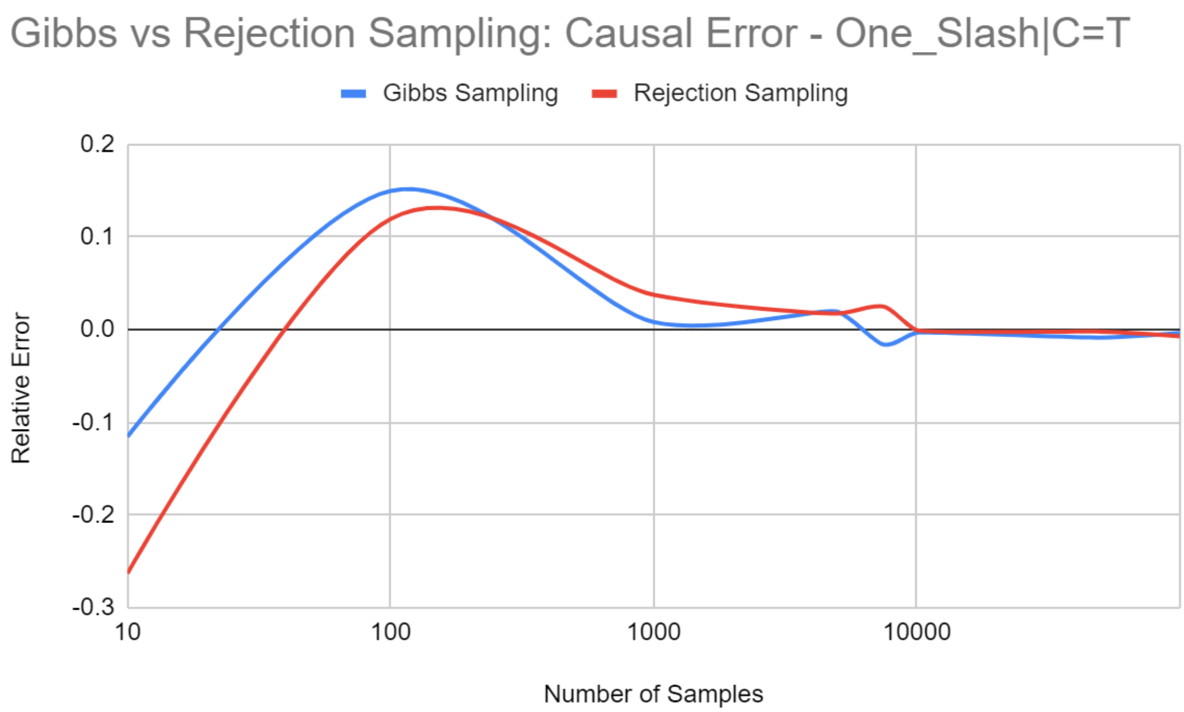
Graph 3: Gibbs and Rejection Sampling Sanity Check Query Convergence for _Close|K=T



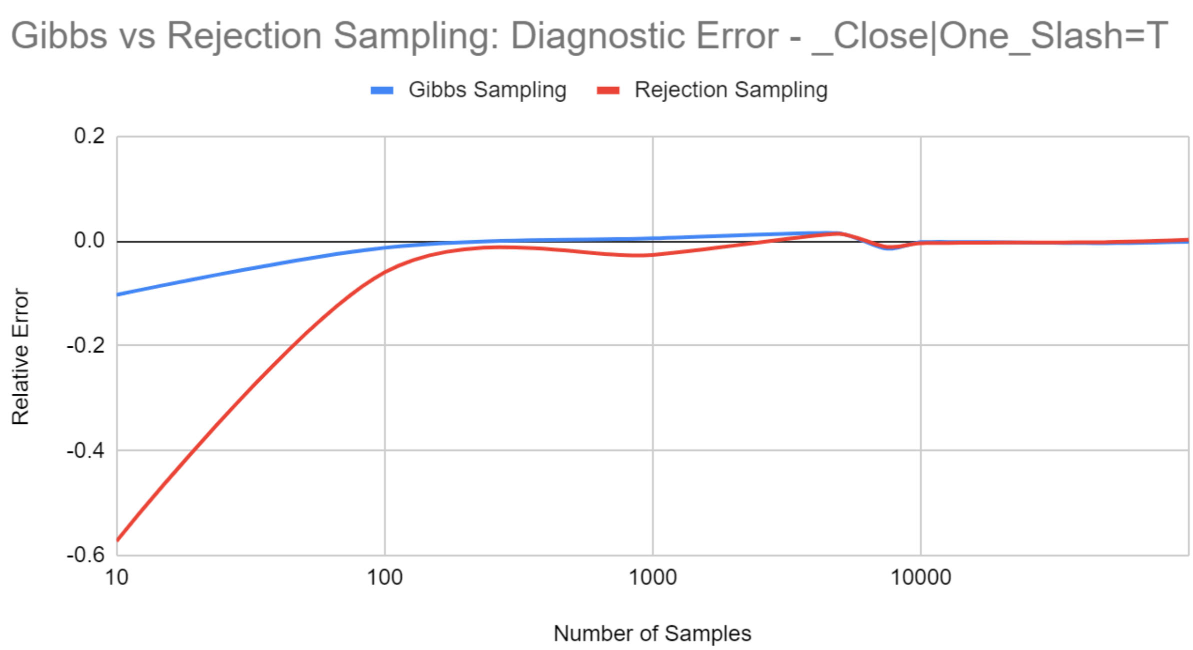
Relative Error Graphs

Below are the relative error graphs for each of the three queries, showing error for both Gibbs and rejection sampling. For the causal and diagnostic queries, the relative error for both Gibbs and rejection sampling approaches zero relative error around 5,000 samples, as seen in Graph 4 and Graph 5. Relative error approaches zero relative error at around 1000 samples for the sanity check query in Graph 6. For this project, the relative error is considered approaching zero relative error when both the tenths and hundredths decimal places are 0.

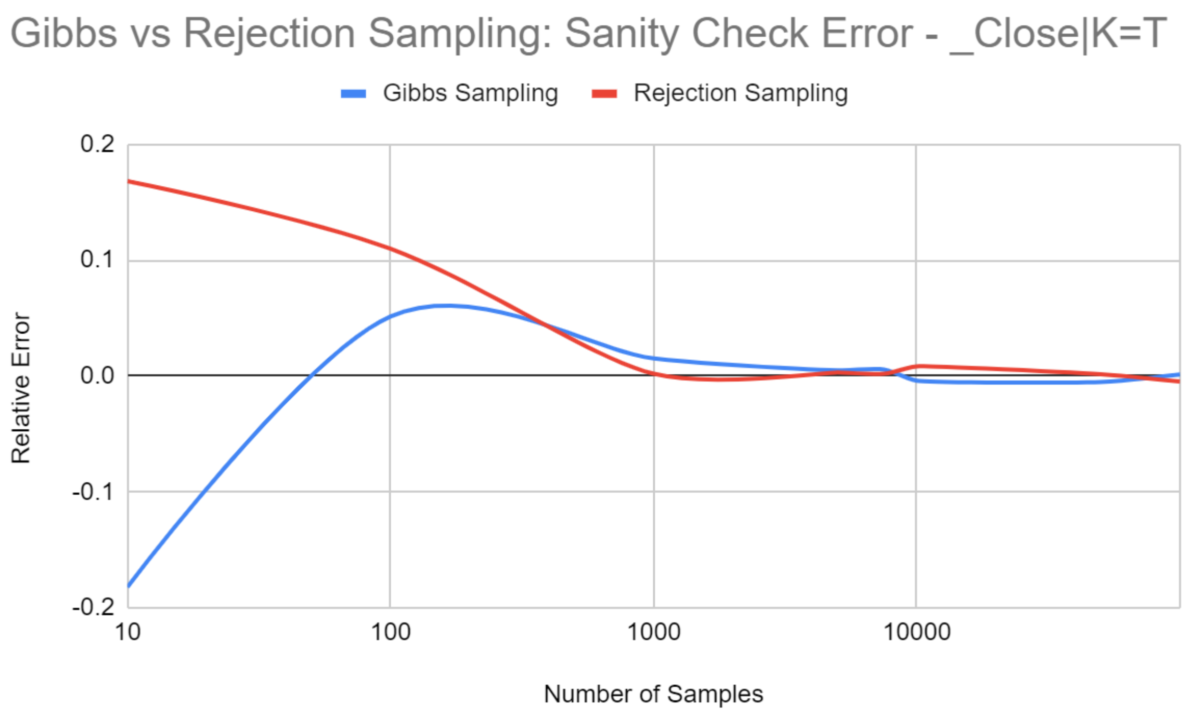
Graph 4: Gibbs and Rejection Sampling Causal Error for One_Slash|C=T



Graph 5: Gibbs and Rejection Sampling Diagnostic Error for _Close|One_Slash=T



Graph 6: Gibbs and Rejection Sampling Sanity Check Error for _Close|K=T



Question and Answer

How did you define stability and accuracy?

Stability can be defined as a "resistance to change, especially sudden change or deterioration"(Dictionary.com 2021). We used this definition as a basis for our requirements for stability. Stability occurred during sampling once we started to consistently get similar values that do not fluctuate very much. There are some fluctuations, but minor ones, because when sampling it is impossible to get the exact same number each time we sample.

Accuracy can be defined as "free from error or defect; consistent with a standard, rule, or model; precise, exact"(Dictionary.com 2021). For this report, we are saying that a probability is accurate if it rounds to the actual probability within 3 to 4 decimal places or if it is within .001 of the actual probability.

How many samples are needed for rejection sampling to become stable? What about to become accurate?

It takes approximately 5,000 samples for rejection sampling to become stable and around 10,000 samples to become accurate.

How many samples are needed for Gibbs sampling to become stable? What about to become accurate?

Gibbs sampling, similar to rejection sampling, generally starts becoming stable around 5,000 samples and accurate around 10,000 samples. Interestingly, it is typically closer to the exact probability at 1,000 samples than rejection sampling is.

Do these values seem to depend on the structure of the network or the query itself or both?

These values seem to depend more on the structure of the network than the query itself. For rejection sampling, stability and accuracy happened in similar places across all the tests. If we apply the same types of queries with the same sample sizes to the AIMA alarm model, it doesn't stabilize at 5,000 or become accurate around 10,000. The query may give small changes for when the values stabilize and become accurate, but for this project, we found the network structure to make the biggest difference.

Is one algorithm "better" than the other?

Both of our algorithms seem to work fairly well and have similar results. We would say that one isn't particularly "better" than the other. Rejection sampling started a little farther away from the actual value than Gibbs, but they both became stable and accurate around the same time. This is visible in the graphs from the previous sections.

Collaboration

Brooke designed, collected data for, calculated probabilities for, and hardcoded the original Bayesian Network implemented in this project. She ran the sampling tests for Gibbs, and helped with implementation of the node representations for the Bayesian Networks. She also designed the algorithms for Gibbs sampling and Markov Blankets. Both Ky and Brooke did troubleshooting and debugging for the entire program. Ky designed the Bayesian Network and node representations, as well as the Enumeration Ask and rejection sampling algorithms. She also hardcoded the AIMA Alarm example for testing, and ran sampling tests for rejection sampling. Writing of the report was split between both team members.

Sources

Dictionary.com. (2021). Definition of accuracy| Dictionary.com. Www.Dictionary.Com. https://www.dictionary.com/browse/accuracy

Dictionary.com. (2021). Definition of stability | Dictionary.com. Www.Dictionary.Com. https://www.dictionary.com/browse/stability

Wiki. (n.d.). *Emma, the Gentle Blade*. Sekiro Shadows Die Twice Wiki. Retrieved November 2021, from https://sekiroshadowsdietwice.wiki.fextralife.com/Emma,+the+Gentle+Blade.