MCMC Lab 3

Lawrence Thatcher
Samantha Sanders

Faculty Data

Iterations:

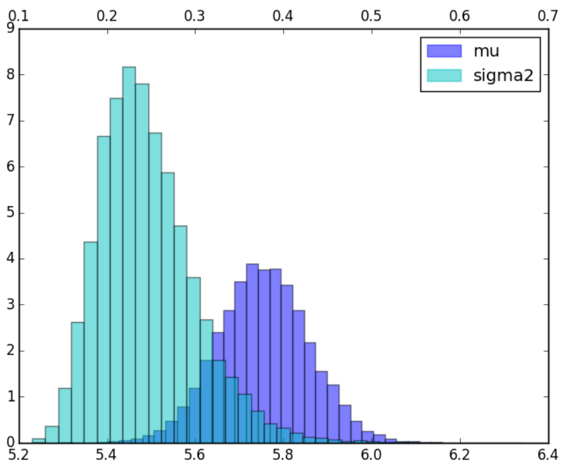

Illustration 1: Mean and Variance without any hyper parameters.

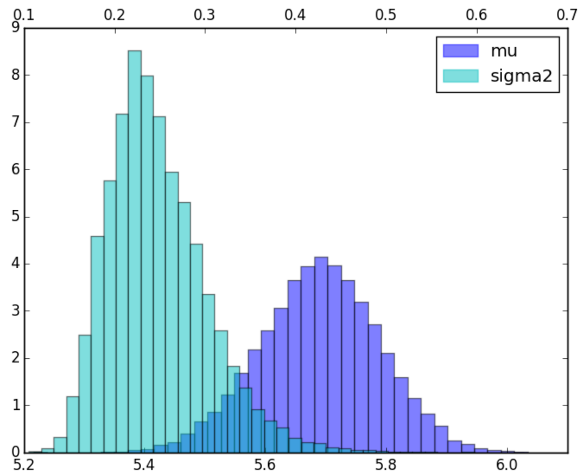

Illustration 2: Mean and Variance with one hyper parameter (mean of μ).

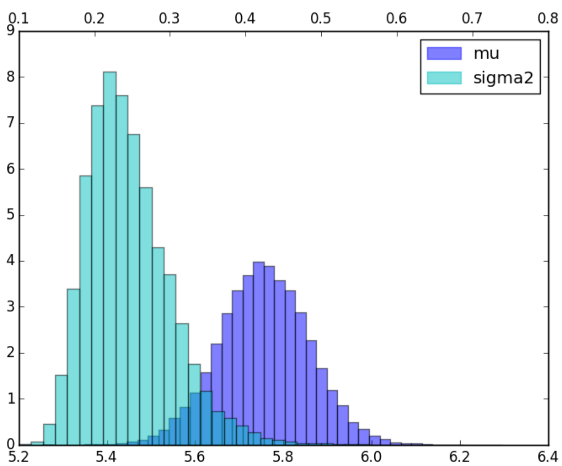

Illustration 4: Mean and Variance with two hyper parameters (mean and variance of μ).

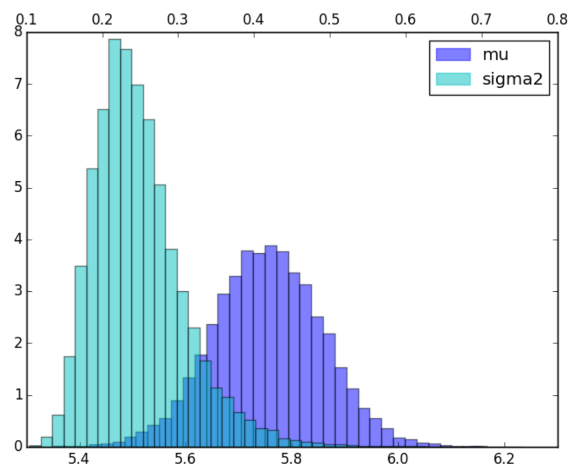

Illustration 3: Mean and Variance with three hyper parameters (μ-μ, μ-σ², σ²-α).

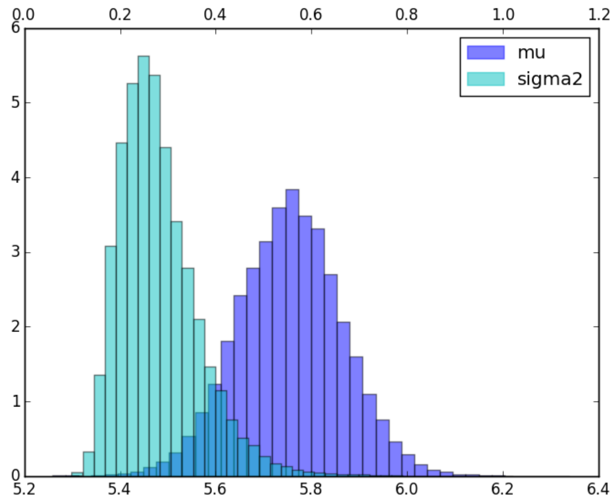

Illustration 5: Mean and Variance with all four hyper parameters.

Posterior of μ-μ over the ages:

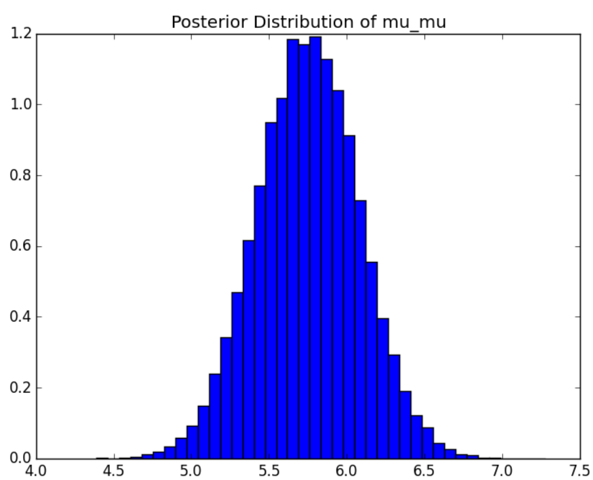

Illustration 6: Posterior distribution of μ-μ with it as the only hyper parameter.

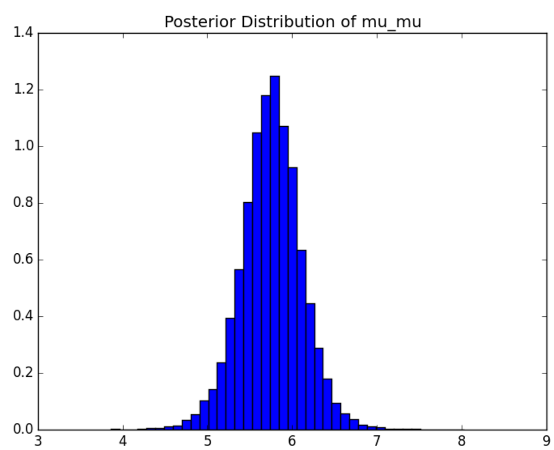

Illustration 7: Posterior distribution of μ-μ with two hyper parameters.

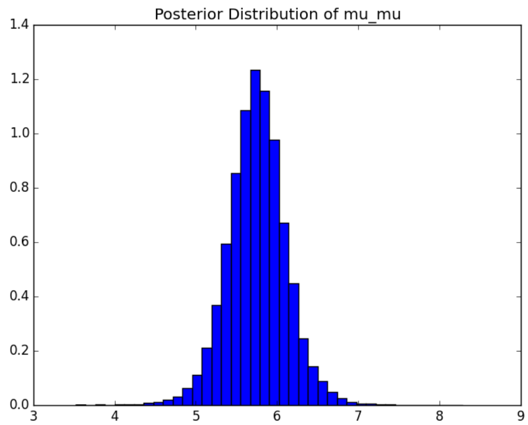

Illustration 8: *Posterior of μ-μ with three hyper parameters.*

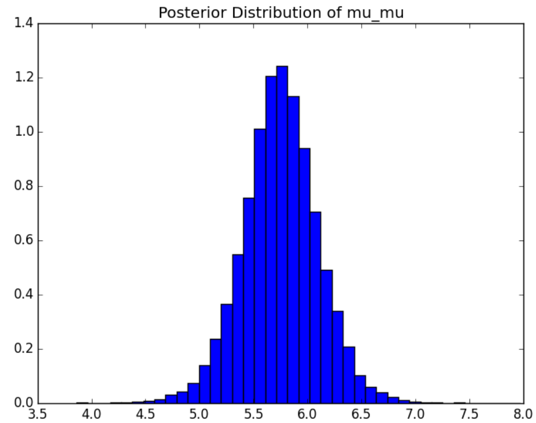

Illustration 9: *Posterior of μ-μ with all four hyper parameters.*

Alarm Networks

Networks

Listed below are the true probabilities of the various network we used in our experiments:

Lab	**Ours**	**Original**
P(B=t): 0.2	P(B=t): 0.2	P(B=t): 0.001
P(E=t): 0.3	P(E=t): 0.7	P(E=t): 0.002
P(A=t | B=t, E=t): 0.95	P(A=t | B=t, E=t): 0.95	P(A=t | B=t, E=t): 0.95
P(A=t | B=t, E=f): 0.94	P(A=t | B=t, E=f): 0.94	P(A=t | B=t, E=f): 0.94
P(A=t | B=f, E=t): 0.29	P(A=t | B=f, E=t): 0.25	P(A=t | B=f, E=t): 0.29
P(A=t | B=f, E=f): 0.2	P(A=t | B=f, E=f): 0.01	P(A=t | B=f, E=f): 0.001
P(J=t | A=t): 0.9	P(J=t | A=t): 0.95	P(J=t | A=t): 0.9
P(J=t | A=f): 0.2	P(J=t | A=f): 0.2	P(J=t | A=f): 0.05
P(M=t | A=t): 0.7	P(M=t | A=t): 0.77	P(M=t | A=t): 0.7
P(M=t | A=f): 0.3	P(M=t | A=f): 0.1	P(M=t | A=f): 0.01

Amount of Data

Shown below is a graph using the networks listed above with various data sizes:

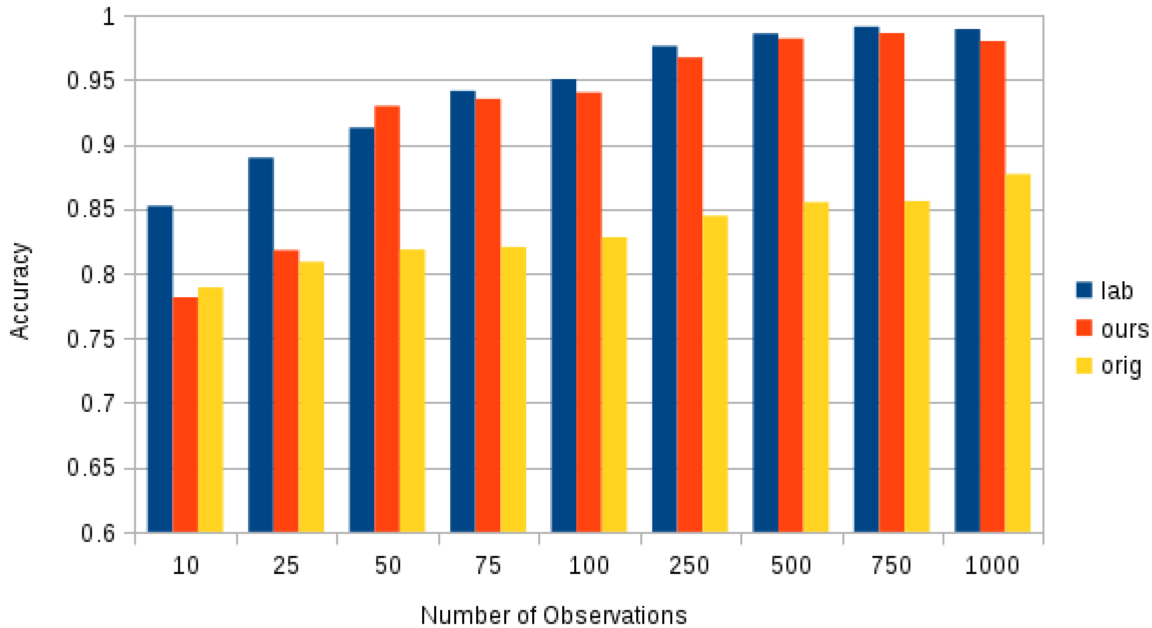

Illustration 10: Accuracy over number of observations for three different networks (specified above).

Note how with our network and the original network, we used smaller values for the true parameters. The smaller the true values are, the more data that will be required to get accurate results.

Number of Hyper-parameters

3 hyper-parameters, 100 observations, 10,000 samples

	Results	Actual
P(B=t)	0.183	0.200
P(E=t)	0.303	0.300
P(M=t | A=t)	0.403	0.700

Comments: The results for B and E are pretty close to the actual values, but P(M=t | A=t) is off by about 0.3.

5 hyper-parameters, 100 observations, 10,000 samples

	Results	Actual
P(B=t)	0.183	0.200
P(E=t)	0.303	0.300
P(M=t | A=t)	0.402	0.700
P(A=t | B=t, E=t)	0.333	0.950
P(A=t | B=f, E=f)	0.334	0.200

Comments: Still good results for B and E, decent results for P(A=t | B=f, E=f), but the rest of the results are pretty inaccurate. Adding more hyper-parameters didn't really change the results for the hyper-parameters from the 3 hyper-parameter experiment.

All hyper-parameters, 100 observations, 10,000 samples

	Results	Actual
P(B=t)	0.183	0.200
P(E=t)	0.303	0.300
P(A=t | B=t, E=t)	0.333	0.950
P(A=t | B=t, E=f)	0.332	0.940
P(A=t | B=f, E=t)	0.331	0.290
P(A=t | B=f, E=f)	0.335	0.200
P(M=t | A=t)	0.400	0.700
P(M=t | A=f)	0.400	0.300
P(J=t | A=t)	0.391	0.900
P(J=t | A=f)	0.391	0.200

Comments: It appears to be easiest to learn the hyper-parameters at the top of the Bayesian network, those ones (B and E) are closest to the actual values.

Hyper-Hyper Parameters

Adding a hyper-hyper-parameter, 100 observations, 50,000 samples

	Results	Actual
P(B=t)	0.184	0.200
E_hyper_alpha	0.622	
P(E=t)	0.301	0.300
P(A=t | B=t, E=t)	0.332	0.950
P(A=t | B=t, E=f)	0.332	0.940
P(A=t | B=f, E=t)	0.331	0.290
P(A=t | B=f, E=f)	0.333	0.200
P(M=t | A=t)	0.401	0.700
P(M=t | A=f)	0.401	0.300
P(J=t | A=t)	0.391	0.900
P(J=t | A=f)	0.392	0.200

Comments: Adding the hyper-hyper parameter to the earthquake node didn't seem to make much of a difference in the results of the non-hyper nodes.

Mixing Plots

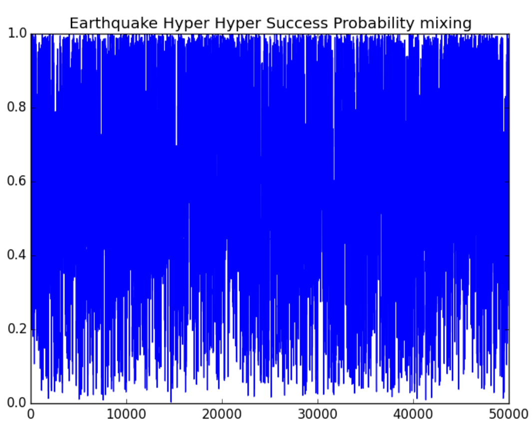

Illustration 11: Mixing Plot of Hyper-Hyper Parameter

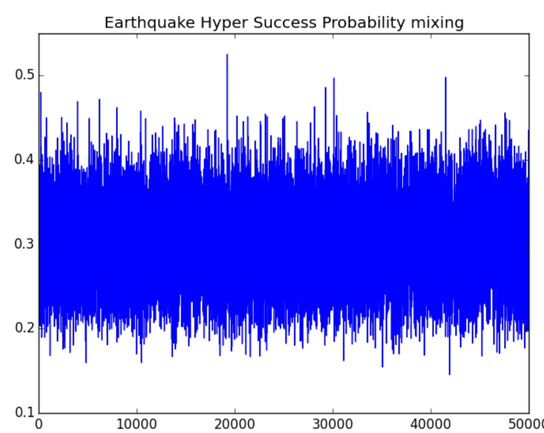

Illustration 12: Mixing Plot of hyper-parameter that has the hyper-hyper parameter.

Missing Data

We also tried our network with various levels of missing data:

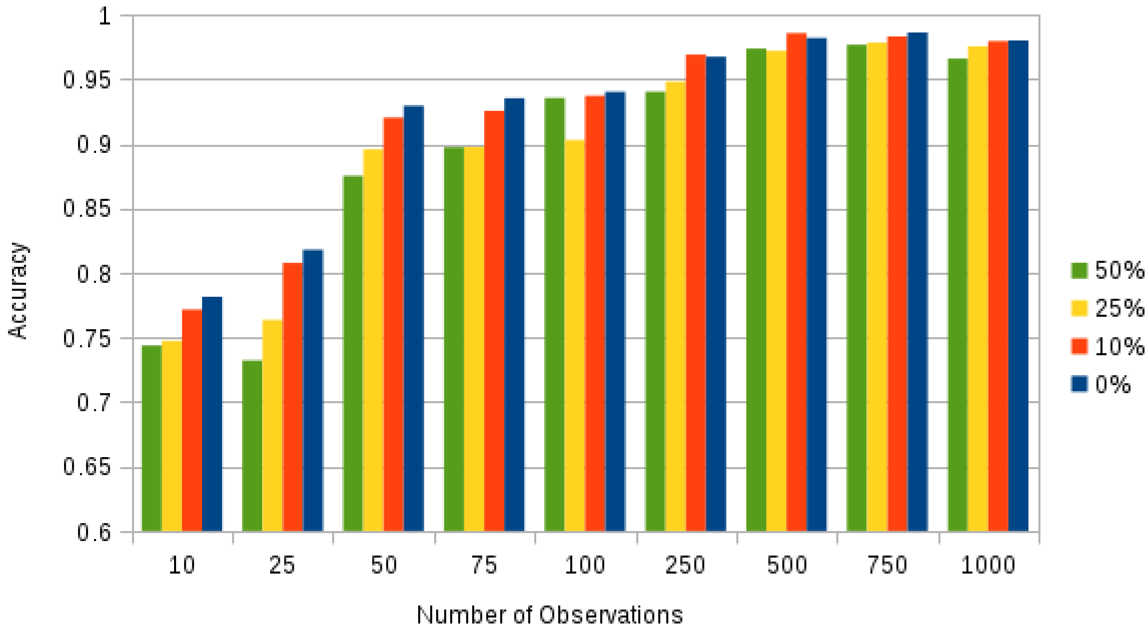

Illustration 13: Accuracy over number of observations for various percentages of missing data on our alarm network.

Inference

We also were able to run inference on different networks:

Lab-Specified Data

	Number of Observations:				
	10	25	50	100	1000
P(B=t | M = t)	0.249	0.2314	0.3489	0.2826	0.2822

Our Data

	Number of Observations:				
	10	25	50	100	1000
P(B=t | M = t)	0.1203	0.1257	0.3039	0.2854	0.375

Posterior Distributions

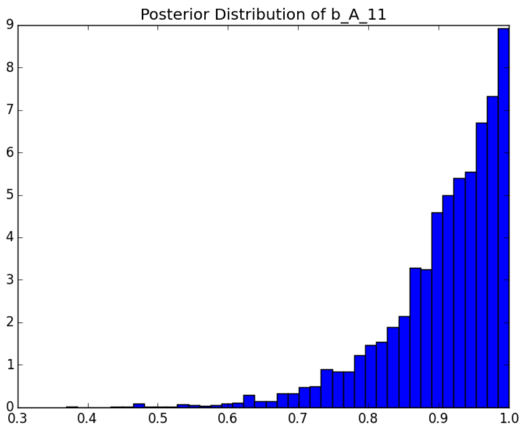

Illustration 14: Posterior distribution for P(A=t | B=t, E=t) on our network.

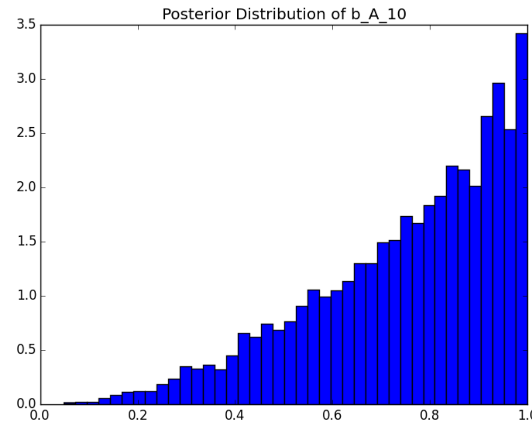

Illustration 15: Posterior distribution for P(A=t | B=t, E=f) on our network.

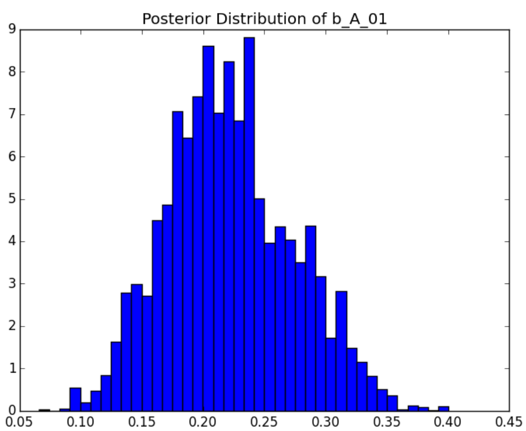

Illustration 17: Posterior distribution for P(A=t | B=f, E=t) on our network.

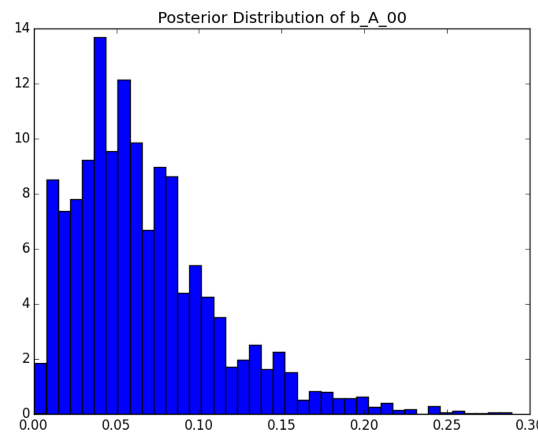

Illustration 16: Posterior distribution for P(A=t | B=f, E=f) on our network.

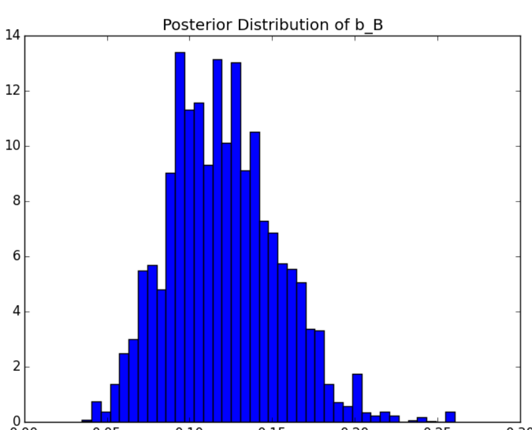

Illustration 18: Posterior distribution for P(B=t) on our network.

Nozomu's Network

Probability Table

	Number of Observations:				
	1	10	50	100	500
P(B=t)	0.670	0.670	0.636	0.618	0.659
P(E=t)	0.668	0.165	0.231	0.207	0.119
P(A=t | B=t, E=t)	0.663	0.671	0.778	0.859	0.926
P(A=t | B=t, E=f)	0.497	0.877	0.963	0.883	0.939
P(A=t | B=f, E=t)	0.501	0.504	0.331	0.300	0.292
P(A=t | B=f, E=f)	0.501	0.604	0.684	0.528	0.487
P(J=t | A=t)	0.665	0.728	0.841	0.883	0.894
P(J=t | A=f)	0.505	0.330	0.196	0.284	0.393
P(M=t | A=t)	0.329	0.548	0.664	0.684	0.677
P(M=t | A=f)	0.498	0.666	0.598	0.432	0.427

Posterior Distributions

Here are the plots of some of the posterior distributions for Nozomu's network:

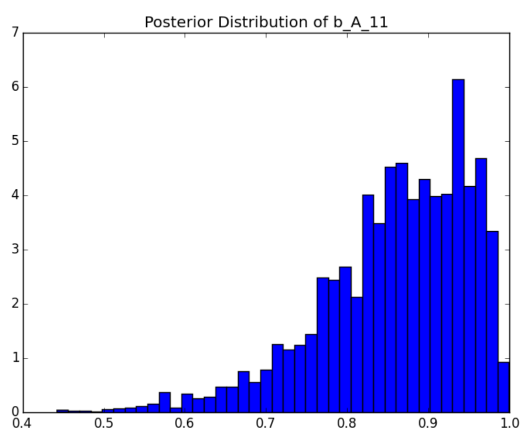

Illustration 19: Posterior distribution for P(A=t | B=t, E=t) on Nozomu's network.

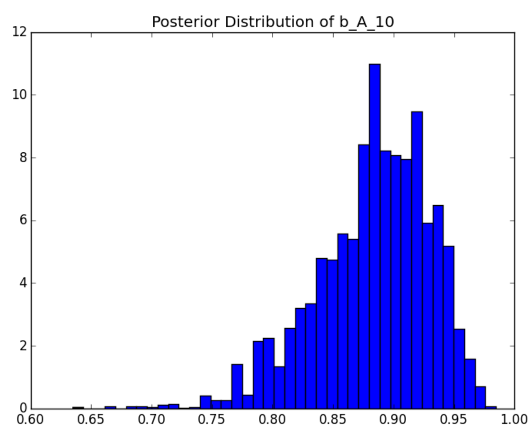

Illustration 20: Posterior distribution for P(A=t | B=t, E=f) on Nozomu's network.

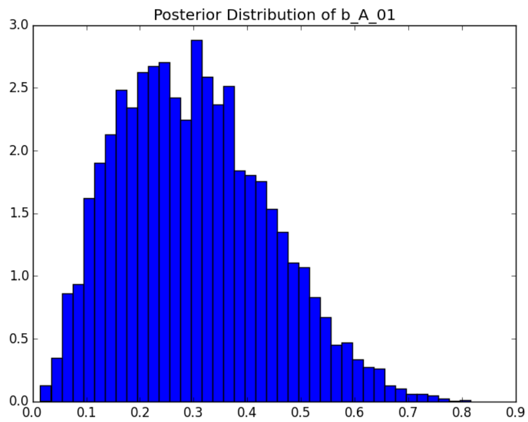

Illustration 21: Posterior distribution for P(A=t | B=f, E=t) on Nozomu's network.

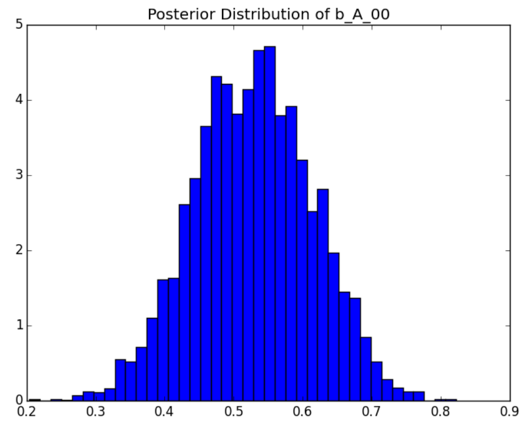

Illustration 22: Posterior distribution for P(A=t | B=f, E=f) on Nozomu's network.

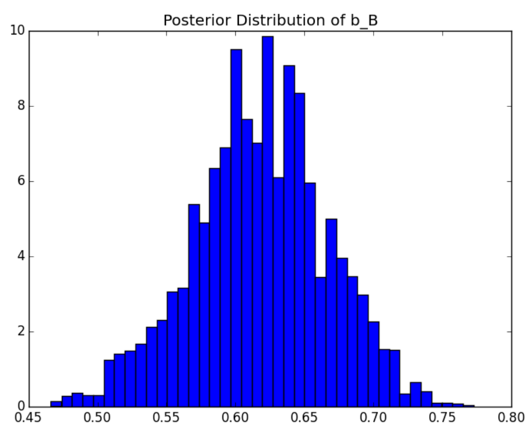

Illustration 23: Posterior distribution for P(B=t) on Nozomu's network.