Y6380872

Bayesian Networks Report

Task 4

Parameter	Posterior
P(0 = 1 | 1 = 0)	0.47891949152542374
P(0 = 1 | 1 = 1)	0.93617021276595747
P(1 = 1)	0.056288742251549694
P(2 = 1 | 0 = 0)	0.30353178607467207
P(2 = 1 | 0 = 1)	0.60982372747078628
P(3 = 1)	0.0092981403719256152
P(4 = 1 | 3 = 0)	0.011200807265388496
P(4 = 1 | 3 = 1)	0.074468085106382975
P(5 = 1 | 1 = 0, 4 = 0)	0.0001072041166380789
P(5 = 1 | 1 = 0, 4 = 1)	0.99122807017543857
P(5 = 1 | 1 = 1, 4 = 0)	0.99821428571428572
P(5 = 1 | 1 = 1, 4 = 1)	0.83333333333333337
P(6 = 1 | 5 = 0)	0.95143653516295024
P(6 = 1 | 5 = 1)	0.020710059171597635
P(7 = 1 | 2 = 0, 5 = 0)	0.10430883213101969
P(7 = 1 | 2 = 0, 5 = 1)	0.70068027210884354
P(7 = 1 | 2 = 1, 5 = 0)	0.79338252796953102
P(7 = 1 | 2 = 1, 5 = 1)	0.921875

These parameter estimates were obtained by using a prior uniform beta distribution. For each occurrence of the parameter in the file, I updated the posterior distribution for that parameter. I then calculated the mean for each parameter at the end.

Task 5

1.
I implemented this sampler in such a way that it can infer the structure of the Bayesian network from the fittedbn parameter.

Firstly, it finds all of the nodes' parents. Then, using this knowledge, it calculates an ordering such that each node's parents occur before the node in the ordering. This is needed because a value cannot be generated for a node without knowledge of its dependencies. After finding a correct order, it iterates nsamples times, and generates a sample, by iterating through the ordering list and using the mean posterior distribution value from fittedbn.

No knowledge of the structure of the graph or the maximum number of parents per node is assumed – any directed acyclic graph will work.

2.
The algorithm is efficient when the maximum number of parents, or nsamples, is small. With respect to maximum number of parents, the time complexity for generating the list is quadratic, and generating the samples is exponential.

3.
nsamples = 1

Parameter	**Posterior**
P(0 = 1 | 1 = 0)	0.33333333333333331
P(0 = 1 | 1 = 1)	0.5
P(1 = 1)	0.33333333333333331
P(2 = 1 | 0 = 0)	0.33333333333333331
P(2 = 1 | 0 = 1)	0.5
P(3 = 1)	0.33333333333333331
P(4 = 1 | 3 = 0)	0.33333333333333331
P(4 = 1 | 3 = 1)	0.5
P(5 = 1 | 1 = 0, 4 = 0)	0.33333333333333331
P(5 = 1 | 1 = 0, 4 = 1)	0.5
P(5 = 1 | 1 = 1, 4 = 0)	0.5
P(5 = 1 | 1 = 1, 4 = 1)	0.5
P(6 = 1 | 5 = 0)	0.66666666666666663
P(6 = 1 | 5 = 1)	0.5
P(7 = 1 | 2 = 0, 5 = 0)	0.33333333333333331
P(7 = 1 | 2 = 0, 5 = 1)	0.5
P(7 = 1 | 2 = 1, 5 = 0)	0.5
P(7 = 1 | 2 = 1, 5 = 1)	0.5

nsamples = 10

Parameter	**Posterior**
P(0 = 1 | 1 = 0)	0.36363636363636365
P(0 = 1 | 1 = 1)	0.66666666666666663
P(1 = 1)	0.16666666666666666
P(2 = 1 | 0 = 0)	0.375
P(2 = 1 | 0 = 1)	0.33333333333333331
P(3 = 1)	0.083333333333333329
P(4 = 1 | 3 = 0)	0.083333333333333329
P(4 = 1 | 3 = 1)	0.5

P(5 = 1 | 1 = 0, 4 = 0)	0.090909090909090912
P(5 = 1 | 1 = 0, 4 = 1)	0.5
P(5 = 1 | 1 = 1, 4 = 0)	0.66666666666666663
P(5 = 1 | 1 = 1, 4 = 1)	0.5
P(6 = 1 | 5 = 0)	0.90909090909090906
P(6 = 1 | 5 = 1)	0.33333333333333331
P(7 = 1 | 2 = 0, 5 = 0)	0.125
P(7 = 1 | 2 = 0, 5 = 1)	0.66666666666666663
P(7 = 1 | 2 = 1, 5 = 0)	0.80000000000000004
P(7 = 1 | 2 = 1, 5 = 1)	0.5

nsamples = 1000

Parameter	**Posterior**
P(0 = 1 | 1 = 0)	0.47284345047923321
P(0 = 1 | 1 = 1)	0.9538461538461539
P(1 = 1)	0.063872255489021951
P(2 = 1 | 0 = 0)	0.2971887550200803
P(2 = 1 | 0 = 1)	0.66403162055335974
P(3 = 1)	0.0099800399201596807
P(4 = 1 | 3 = 0)	0.015105740181268883
P(4 = 1 | 3 = 1)	0.18181818181818182
P(5 = 1 | 1 = 0, 4 = 0)	0.0010822510822510823
P(5 = 1 | 1 = 0, 4 = 1)	0.94117647058823528
P(5 = 1 | 1 = 1, 4 = 0)	0.98461538461538467
P(5 = 1 | 1 = 1, 4 = 1)	0.5
P(6 = 1 | 5 = 0)	0.94372294372294374
P(6 = 1 | 5 = 1)	0.037499999999999999
P(7 = 1 | 2 = 0, 5 = 0)	0.12955465587044535
P(7 = 1 | 2 = 0, 5 = 1)	0.7857142857142857
P(7 = 1 | 2 = 1, 5 = 0)	0.78240740740740744
P(7 = 1 | 2 = 1, 5 = 1)	0.90740740740740744

nsamples = 1000000

Parameter	**Posterior**
P(0 = 1 | 1 = 0)	0.47866364265216382
P(0 = 1 | 1 = 1)	0.93705919114364788

Y6380872

P(1 = 1)	0.056003887992224013
P(2 = 1 | 0 = 0)	0.30401923876158543
P(2 = 1 | 0 = 1)	0.60918867902081542
P(3 = 1)	0.0092889814220371552
P(4 = 1 | 3 = 0)	0.011305987398986993
P(4 = 1 | 3 = 1)	0.076103336921420878
P(5 = 1 | 1 = 0, 4 = 0)	0.0001029233446227377
P(5 = 1 | 1 = 0, 4 = 1)	0.99219027334043308
P(5 = 1 | 1 = 1, 4 = 0)	0.99813961889280234
P(5 = 1 | 1 = 1, 4 = 1)	0.84890965732087231
P(6 = 1 | 5 = 0)	0.95113423079932646
P(6 = 1 | 5 = 1)	0.021451677822334193
P(7 = 1 | 2 = 0, 5 = 0)	0.10402065011200425
P(7 = 1 | 2 = 0, 5 = 1)	0.69632933104631223
P(7 = 1 | 2 = 1, 5 = 0)	0.79356330268239372
P(7 = 1 | 2 = 1, 5 = 1)	0.92304853293965672

When the number of samples is 1, the accuracy is very bad, and many parameters are off by at least 0.5.
With 10 samples, the parameters are usually within 0.5 of the mean. With 1,000, parameters that are produced a large number of times are around 0.01 away from the mean, however there are still some parameters, such as P(5 = 1 | 1 = 1, 4 = 1), that have not been generated yet, since the condition is unlikely. With 1,000,000 samples, most parameters are correct to three decimal places.
As the number of samples tends to infinity, the sample results will approach the mean of the parameters.