Assignment 4

Maxime CHAMBREUIL
McGill ID: 260067572
maxime.chambreuil@mail.mcgill.ca

Contents

1 Exercise 1 : Likelihood weighting

Let's take an ordering of the variables consistent with the arc direction in the Bayes net structure: A, B, C, D, E.

To generate the first sample, we assume that w = 1.

- A has not been observed so we can sample it directly: A = 1
- B has not been observed and as p (B = 1 / A = 1) = 0,8, we can sample B = 1
- C has been observed: C = 1 and the new w = p (C = 1 / A = 1) = 0,6
- D has been observed: D = 1 and the new w = w $\times$ p (D = 1 / B = 1 , C = 1) = $0,6 \times 1$ = 0,6
- E has not been observed and as p (E = 1 / D = 1) = 0,8, we can sample E = 1

We finally obtained our first sample: (A = 1 , B = 1 , C = 1 , D = 1 , E = 1) with a w = 0,6.

To generate the second sample, we assume that w = 1.

- A has not been observed so we can sample it directly: A = 0
- B has not been observed and as p (B = 0 / A = 0) = 0,8, we can sample B = 0
- C has been observed: C = 1 and the new w = p (C = 1 / A = 0) = 0,3
- D has been observed: D = 1 and the new w = w $\times$ p (D = 1 / B = 0 , C = 1) = $0,3 \times 0 = 0$
- E has not been observed and as p (E = 0 / D = 1) = 0,8, we can sample E = 0

We finally obtained our second sample: (A = 0 , B = 0 , C = 1 , D = 1 , E = 0) with a w = 0.

The likelihood weighting will encounter a problem in this network: as we have seen in the second sample, we have a sample with zero as a weight, which is not really useful. This problem is due to the deterministic AND in the V-structure.

2 Exercise 2 : Markov Chain

2.1 From 0 to 0 in n steps

To go from state 0 to state 0 in n steps, we have to keep staying in this state, as we cannot go to state 0 from 1 or 2.

Therefore, we have:

$$p_{00}^{(n)} = \left(\frac{1}{2}\right)^n$$

2.2 From 0 to 1 in n steps

To go from state 0 to state 1 in n steps, we can stay in state 0 during m steps, then we can move to state 1, then move to state 2, stay there p steps and come back to state 1 q times, or stay until the end in state 1.

Therefore, we have:

$$p_{01}^{(n)} = \left(\frac{1}{2}\right)^m \times \frac{1}{4} \times \left(\frac{1}{2}\right)^q \times \left(\frac{1}{2}\right)^p \times \left(\frac{1}{2}\right)^q \times \left(\frac{1}{2}\right)^{n-m-1-p-2q} = \left(\frac{1}{2}\right)^{n+1}$$

You can obtain $p_{02}^{(n)} = \left(\frac{1}{2}\right)^{n+1}$ with the same method.

2.3 From 1 or 2 to 0

As we cannot go to state 0 from any other states, $p_{10}^{(n)} = p_{20}^{(n)} = 0$

2.4 From 1 to 1 in n steps

To go from state 1 to state 1 in n steps, we can stay in state 1 during m steps, then we can move to state 2, stay there p steps and come back to state 1.

Therefore, we have:

$$p_{11}^{(n)} = \left(\frac{1}{2}\right)^m \times \left(\frac{1}{2}\right)^{\frac{n-m-p}{2}} \times \left(\frac{1}{2}\right)^p \times \left(\frac{1}{2}\right)^{\frac{n-m-p}{2}} = \left(\frac{1}{2}\right)^n$$

You can obtain $p_{22}^{(n)} = \left(\frac{1}{2}\right)^n$ with the same method.

2.5 From 1 to 2 in n steps

To go from state 1 to state 2 in n steps, we can stay in state 1 during m steps, then we can do p loops (state 1, state 2, state 1), stay in state 2 until n-th steps.

Therefore, we have:

$$p_{12}^{(n)} = \left(\frac{1}{2}\right)^m \times \left(\frac{1}{2}\right)^{2p} \times \left(\frac{1}{2}\right)^{n-m-2p} = \left(\frac{1}{2}\right)^n$$

You can obtain $p_{21}^{(n)} = \left(\frac{1}{2}\right)^n$ with the same method.

2.6 Transition probability matrix

$$\begin{bmatrix} \left(\frac{1}{2}\right)^n & \left(\frac{1}{2}\right)^{n+1} & \left(\frac{1}{2}\right)^{n+1} \\ 0 & \left(\frac{1}{2}\right)^n & \left(\frac{1}{2}\right)^n \\ 0 & \left(\frac{1}{2}\right)^n & \left(\frac{1}{2}\right)^n \end{bmatrix}$$

3 Exercise 3 : Gibbs sampling

4 Exercise 4 : Gibbs sampling

4.1 With deterministic exclusive OR

To show that Gibbs sampling on the structure with evidence Z = 1 will estimate p (X = 1 / Z = 1) as either 1 or 0, let us take a look at the application of the algorithm:

To initialize, we can pick a random sample (X = 0 , Y = 1 , Z = 1) and repeat the following procedure:

- Pick a non-evidence variable: X
- Sample x from p (x / y , z): p (X = 1 / Y = 1 , Z = 1) = 0 so we obtain a new sample (X = 1 , Y = 1 , Z = 1)

we can also do it with Y:

- Pick a non-evidence variable: Y
- Sample x from p (y / x , z): p (Y = 0 / X = 1 , Z = 1) = 1 so we obtain a new sample (X = 1 , Y = 0 , Z = 1)

Let's do it another time with X:

- Pick a non-evidence variable: X
- Sample x from p (x / y , z): p (X = 1 / Y = 0 , Z = 1) = 1 so we obtain the previous sample.

You can notice that Gibbs sampling on the structure with evidence Z = 1 estimates p (X = 1 / Z = 1) as either 1 or 0.

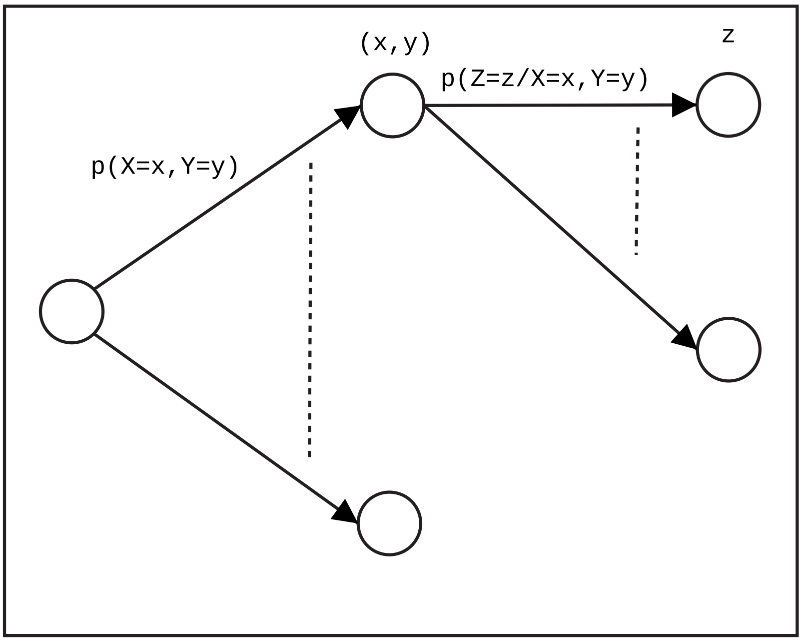
4.2 With slightly noisy exclusive OR

We can repeat the previous algorithm with the same sample, and the change would be in the computation of the probability.

5 Exercise 5 : Parameter estimation in Bayes nets

Let us consider a simple V-structure: $X \rightarrow Z \leftarrow Y$, where X can take n different values, Y p values and Z q values.

Given x and y, Z can take any one of his q possible values with a probability distribution depending on x and y.



With this representation, it is easy to see that each probability distribution of p(Z/X,Y) is independent with the other.

6 Exercise 6 : Maximum Likelihood Estimation

6.1 Likelihood of λ

$$\begin{aligned} L(\lambda/D) &= \prod_{j=1}^{m} p(x_j/\lambda) \\ &= \prod_{j=1}^{m} e^{-\lambda} \frac{\lambda^{x_j}}{x_j!} \end{aligned}$$

$$= \quad e^{-m\lambda} \prod_{j=1}^{m} \frac{\lambda^{x_j}}{x_j!}$$

6.2 Derivation of the likelihood

$$\begin{aligned} logL(\lambda/D) &= -m\lambda + \sum_{j=1}^{m} log\left(\frac{\lambda^{x_j}}{x_j!}\right) \\ \frac{\partial\left[logL(\lambda/D)\right]}{\partial\lambda} &= -m + \sum_{j=1}^{m} \frac{x_j}{\lambda} = 0 \Rightarrow \lambda = \frac{\sum_{j=1}^{m} x_j}{m} \end{aligned}$$

We have obtained the average of x_j, so the sufficient statistic of the data in this case is the mean of the data.

6.3 Relation between λ_n and λ_{n+1}

$$\begin{aligned} \lambda_{n+1} &= \frac{1}{n+1} \sum_{j=1}^{n+1} x_j \\ &= \frac{1}{n+1} \sum_{j=1}^{n} x_j + \frac{1}{n+1} \times x_{n+1} \\ &= \frac{n}{n+1} \times \frac{1}{n} \sum_{j=1}^{n} x_j + \frac{1}{n+1} \times x_{n+1} \\ &= \frac{n}{n+1} \lambda_n + \frac{x_{n+1}}{n+1} \\ &= \frac{n\lambda_n + x_{n+1}}{n+1} \end{aligned}$$

7 Exercise 7 : Learning Bayes net structure

7.1 Relationship between B_1 and B_2

If we have an additional arc in B_2 compared to B_1, it means that there will be an extra term in the computation of the likelihood of the network. So the likelihood of B_2 will be greater than the one of B_1.

7.2 Consequences

It means that more we add arc, better the network will be scored. So we will obtain a fully connected graph, which will overfit the data and will be really bad with new samples.