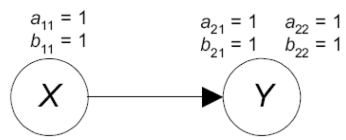
Probabilistic Reasoning

Unit # 13

Learning Parameters

- For each probability in the network there is a pair (aij, bij). The i indexes the variable; the j indexes the value of the parent(s) of the variable.
- For example, the pair (a11, b11) is for the first variable (X) and the first value of its parent (in this case there is a default of one parent value since X has no parent).
- The pair (a21, b21) is for the second variable (Y) and the first value of its parent, namely x1.
- We have attempted to represent prior ignorance as to the value of all probabilities by taking aij = bij = 1.



$P(x_1) = a_{11} / (a_{11}+b_{11}) = 1/2$

$P(y_1 | x_1) = a_{21} / (a_{21}+b_{21}) = 1/2$

$P(y_1 | x_2) = a_{22} / (a_{22}+b_{22}) = 1/2$

New Data

- When we obtain data, we use an (sij, tij) pair to represent the counts for the ith variable when the variable's parents have their jth value.

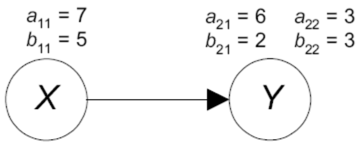
Case	X	Y
1	x_1	y_1
2	x_1	y_1
3	x_1	y_1
4	x_1	y_1
5	x_1	y_1
6	x_1	y_2
7	x_2	y_1
8	x_2	y_1
9	x_2	y_2
10	x_2	y_2

$s_{11} = 6$ $t_{11} = 4$

$s_{21} = 5$ $t_{21} = 1$ $s_{22} = 2$ $t_{22} = 2$

Updated Probabilities

- To determine the posterior probability distribution based on the data, we update each conditional probability with the counts relative to that conditional probability. Since we want an updated Bayesian network, we re-compute the values of the (aij, bij) pairs.

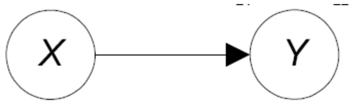


$P(x_1) = a_{11} / (a_{11}+b_{11}) = 7/12$

$P(y_1 | x_1) = a_{21} / (a_{21}+b_{21}) = 3/4$

$P(y_1 | x_2) = a_{22} / (a_{22}+b_{22}) = 1/2$

Right Prior?



- Should we assume a prior distribution?
- If yes, what values should be considered for the prior distribution?

Case	X	Y
1	x_1	y_1
2	x_1	y_1
3	x_1	y_1
4	x_1	y_1
5	x_1	y_1
6	x_1	y_2
7	x_2	y_1
8	x_2	y_1
9	x_2	y_2
10	x_2	y_2

Equivalent Sample Size for Prior

- First try
  - N(x1) = N(x2) = N(y1|x1) = N(y2|x1) = N(y1|x2) = N(y2|x2) = 1
  - where N(a) is the number of cases
  - Compute P(y1). Is it consistent with the prior?
- Then
  - Use the following data set
  - Now compute P(y1).

Case	X	Y
1	x_1	y_1
2	x_1	y_2
3	x_2	y_1
4	x_2	y_2

Theorem

- Suppose we specify a Bayesian network for parameter learning in the case of binomial variables and assign for all i and j

$$a_{ij} = b_{ij} = N / 2q_i$$

- where N is a positive integer and q_i is the number of instantiations of the parents of the ith variable. Then the resultant Bayesian network has equivalent sample size N , and the joint probability distribution in the Bayesian network is uniform.

Example (Source: Neapolitan)

- Assume that you feel your prior experience concerning the relative frequency of smokers in a particular bar is equivalent to having seen 14 smokers and 6 nonsmokers.
  - You then decide to poll individuals in the bar and ask them if they smoke. What is your probability of the first individual you poll being a smoker?
  - Suppose that after polling 10 individuals, you obtain these data (the value 1 means the individual smokes and 2 means the individual does not smoke):

    {1, 2, 2, 2, 2, 1, 2, 2, 2, 1}

    What is your probability that the next individual you poll is a smoker?

Example (Cont'd)

– Suppose that after polling 1000 individuals (it is a big bar), you learn that 312 are smokers. What is your probability that the next individual you poll is a smoker? How does this probability compare to your prior probability?

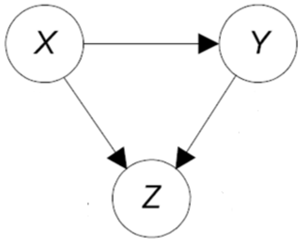
Example II (Source: Neapolitan)

- Suppose that you are going to sample individuals who have smoked two packs of cigarettes or more daily for the past 10 years. You will determine whether each individual's systolic blood pressure is $\leq$100, 101-120, 121-140, 141-160, or $\geq$161. Determine values of a1, a2, …, a5 that represent your prior probability of each blood pressure range.
- Next you sample such smokers. What is your probability of each blood pressure range for the first individual sampled?
- Suppose that after sampling 100 individuals, you obtain the following results:
- Compute your probability of each range for the next individual sampled.

Blood Pressure Range	# of Individuals in This Range
≤ 100	2
101-120	15
121-140	23
141-160	25
≥ 161	35

Example III (Source: Neapolitan)

- Suppose that we have the following Bayesian network for parameter learning and the following data.
- Determine the updated BN for parameter learning.



Case	X	Y	Z
1	x_1	y_2	z_1
2	x_1	y_1	z_2
3	x_2	y_1	z_1
4	x_2	y_2	z_1
5	x_1	y_2	z_1
6	x_2	y_2	z_2
7	x_1	y_2	z_1
8	x_2	y_1	z_2
9	x_1	y_2	z_1
10	x_1	y_1	z_1
11	x_1	y_2	z_1
12	x_2	y_1	z_2
13	x_1	y_2	z_1
14	x_2	y_2	z_2
15	x_1	y_2	z_1

Homework Problem

- Develop Bayesian networks for parameter learning with equivalent sample sizes 1, 2, 4, and 8 for the following DAG.