PROBABILISTIC REASONING

Unit # 11

ACKNOWLEDGEMENT

The material in this presentation is taken from Richard Neapolitan's book "Probabilistic Method for Bioinformatics: With an Introduction to Bayesian Networks"

LEARNING IN BN

Until the early 1990s the DAG in a Bayesian network was ordinarily hand-constructed by a domain expert.

Then the conditional probabilities were assessed by the expert, learned from data, or obtained using a combination of both techniques.

Eliciting Bayesian networks from experts can be a laborious and difficult process in the case of large networks.

As a result, researchers developed methods that could learn the DAG from data.

Furthermore, they formalized methods for learning the conditional probabilities from data.

MLE VS. MAP

Suppose you are going to toss a coin.

If you toss it 100 times and get 48 heads then the *maximum likelihood estimate (MLE)* is

- P(heads) ≈ 0.48

However, if you have a prior belief the relative frequency is around = 0.5, you might feel your prior experience is equivalent to having seen 50 heads in 100 tosses. In such case, your *maximum a posterior probability (MAP)* is

- P(heads | 48, 52) = 98/200 = 0.49

HOW TO ESTIMATE PROBABILITIES FROM DATA?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Class: $P(C) = N_c/N$

- $P(No) =$?
  $P(Yes) =$?

For discrete attributes:

$$P(A_i \mid C_k) = |A_{ik}| / N_{Ck}$$

- where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
- Examples:

  P(Status=Married|No) = ?

  P(Refund=Yes|Yes)= ?

HOW TO ESTIMATE PROBABILITIES FROM DATA?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Class: $P(C) = N_c/N$

- e.g., $P(No) = 7/10$, $P(Yes) = 3/10$

For discrete attributes:

$$P(A_i \mid C_k) = |A_{ik}| / N_{Ck}$$

- where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
- Examples:

  $P(Status=Married|No) = 4/7$
  $P(Refund=Yes|Yes)=0$

LAPLACIAN SMOOTHING

When we estimate probabilities from data, it can sometimes happen that a particular event never occurs in the dataset.

To avoid having probability value being zero, implying an *impossible* case; we use some type of *smoothing* for the probabilities, eliminating zero probability values.

There are several smoothing techniques, one of the most common and simplest being *Laplacian* smoothing.

Laplacian smoothing consists in initializing the probabilities to a uniform distribution, and then updating these values based on the data.

LAPLACIAN SMOOTHING

Laplacian smoothing consists in initializing the probabilities to a uniform distribution, and then updating these values based on the data. Consider a discrete variable, X, *with* k *possible values. Initially, each probability will be set to* $P(x_i) = 1/k$.

Then, consider a dataset with N *samples, in which the value* x_i occurs m *times; the* estimate of its probability will be the following:

$$P(x_i)) = (1 + m)/(k + N)$$

THUMBTACK EXAMPLE

Suppose you are going to repeatedly toss a thumbtack.

Based on its structure, you might feel it should land heads about half the time, but you are not nearly so confident as you were with the coin from your pocket.

So, you might feel your prior experience is equivalent to having seen 3 heads in 6 tosses.

Then your prior probability of heads is

- P(head) = 3/6 = 0.5

After seeing 65 heads in 100 tosses, your posterior probability is

- P(head | 65, 35) = 68 / 106 = 0.64

DICE EXAMPLE

Suppose we have an asymmetrical-, six-sided die, and we have little idea of the probability of each side coming up. However, it seems that all sides are equally likely. So, we assign 3 to each outcome.

Suppose next we throw the die 100 times, with the following results:

Outcome	Number of Occurrences
1	10
2	15
3	5
4	30
5	13
6	27

DICE EXAMPLE (CONTD.)

$$
\begin{aligned}
P(1|10,15,5,30,13,27) &= \frac{a_1+s_1}{m+n} = \frac{3+10}{18+100} = .110 \\
P(2|10,15,5,30,13,27) &= \frac{a_2+s_2}{m+n} = \frac{3+15}{18+100} = .153 \\
P(3|10,15,5,30,13,27) &= \frac{a_3+s_3}{m+n} = \frac{3+5}{18+100} = .067 \\
P(4|10,15,5,30,13,27) &= \frac{a_4+s_4}{m+n} = \frac{3+30}{18+100} = .280 \\
P(5|10,15,5,30,13,27) &= \frac{a_5+s_5}{m+n} = \frac{3+13}{18+100} = .136 \\
P(6|10,15,5,30,13,27) &= \frac{a_6+s_6}{m+n} = \frac{3+27}{18+100} = .254.
\end{aligned}
$$

CONFIDENCE IN PRIOR EXPERIENCE

Complete ignorance: $a = b = 0$.

Low confidence: $a + b$ *small* (10).

Medium confidence: $a + b$ *intermediate* (100).

High confidence: $a + b$ large (1000).

EXAMPLE

For example, let us assume an expert gives an estimate of 0.7 for a certain parameter, and that the experimental data provides 40 positive cases among 100 samples. The parameter estimation for different confidences assigned to the expert will be the following:

Low confidence ($a + b = 10$): $P(b_i) = \frac{40+7+1}{100+10+2} = 0.43$

Medium confidence ($a + b = 100$): $P(b_i) = \frac{40+70+1}{100+100+2} = 0.55$

High confidence ($a + b = 1000$): $P(b_i) = \frac{40+700+1}{100+1000+2} = 0.67$

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?

EXERCISE

Example 7.4 *Suppose you are going to repeatedly toss a thumbtack. Based on its structure, you might feel it should land heads about half the time, but you are not nearly so confident as you were with the coin from your pocket. So, you might feel your prior experience is equivalent to having seen 3 heads in 6 tosses. Then your prior probability of heads is*

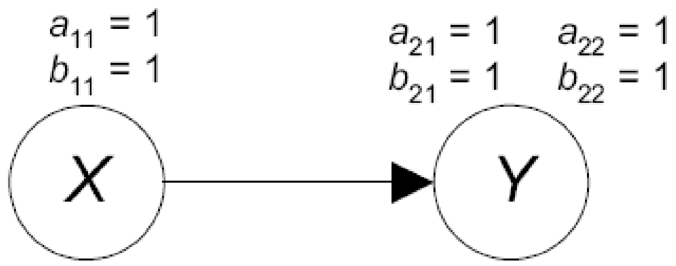
LEARNING PARAMETERS

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.

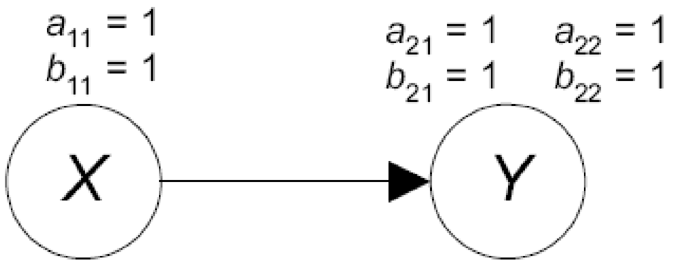
We have attempted to represent prior ignorance as to the value of all probabilities by taking aij = bij = 1.



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

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

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



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

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

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

$$P(x_1) = \frac{1}{2}$$

$$P(x_2) = \frac{1}{2}$$

$\cancel{P(y}$ $P(Y|X)$

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

$P(Y|X)$

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

$\bar{Y}$

$$P(y_2|x_1) = \frac{b_{21}}{a_{21}+b_{21}}$$

$$P(y_2|x_2) = \frac{b_{22}}{a_{22}+b_{22}}$$

NEW DATA

$a_{11}' = a_{11} + s_{11} = 1+6=7$

$b_{11}' = b_{11} + t_{11} = 1+4=5$

$b_{22}' = b_{22} + t_{22} = 1+2=3$

$a_{22}' = a_{22} + s_{22} = 1+2=3$

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.

Prior	Data		
a_{11}	s_{11}	a_{22}	s_{22}
b_{11}	t_{11}	b_{22}	t_{22}
a_{21}	s_{21}		
b_{21}	t_{21}		

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

Cases 1–6: $s_{11} = 6$

Cases 7–10: $t_{11} = 4$

Cases 1–5: $s_{21} = 5$

Case 6: $t_{21} = 1$

Cases 7–8: $s_{22} = 2$

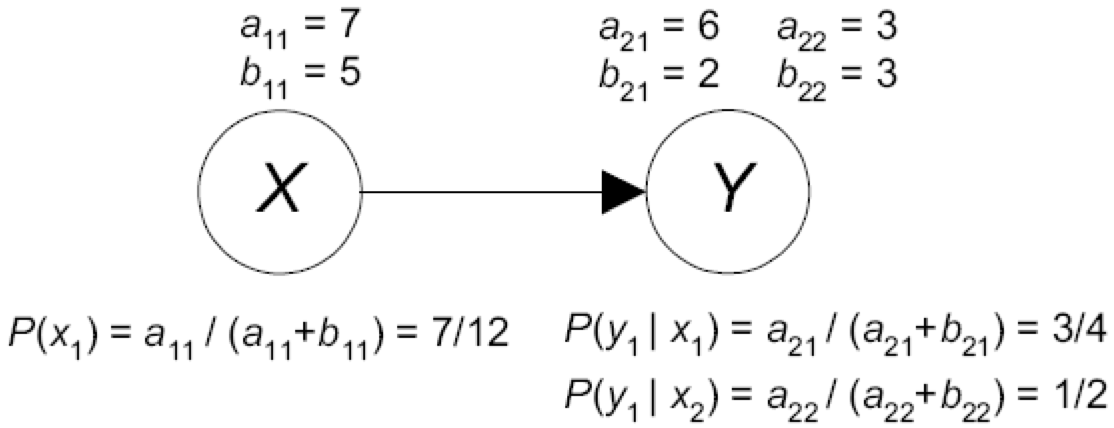
Cases 9–10: $t_{22} = 2$

$$P(x_1) = \frac{a_{11}'}{a_{11}' + b_{11}'} = \frac{7}{12}$$

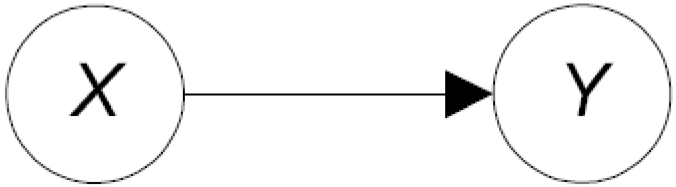
$$P(y_2 \mid x_2) = \frac{b_{22}'}{b_{22}' + a_{22}'} = \frac{3}{3+3} = \frac{1}{2}$$

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.



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

Then

- Use the following data set

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

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.

EXAMPLE II (CONT'D)

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:

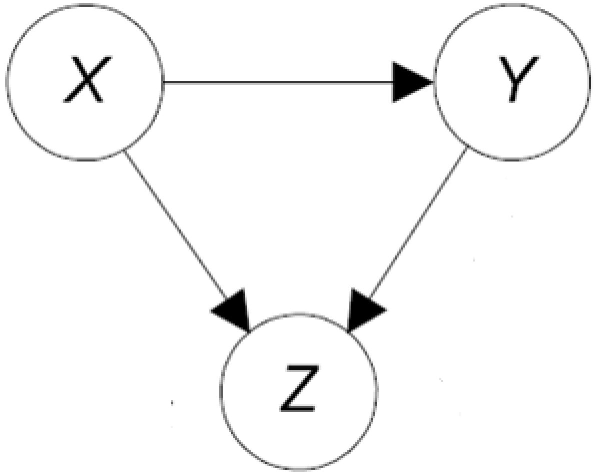
Compute your probability of each range for the next individual sampled.

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

$X \quad Y \quad Z$

$x_1 \quad y_1 \quad z_1$

$x_1 \quad y_1 \quad z_2$

$x_1 \quad y_2 \quad z_1$

$x_1 \quad y_2 \quad z_2$

$x_2 \quad y_1 \quad z_1$

$x_2 \quad y_1 \quad z_2$

$x_2 \quad y_2 \quad z_1$

$x_2 \quad y_2 \quad z_2$