Assignment2 – Bayesian Networks

Given: May 2   Due: May 16

Problem 2.1 (Is your TA in the office?)

You want to discuss something with your TA. You know that

1. the probability of your TA being in the office, assuming it is morning, is $\frac{1}{5}$,
2. if your TA is in the office, there is a $\frac{1}{3}$ probability it is morning,
3. the probabilities that it is morning or afternoon are both $\frac{1}{2}$

Your tasks:

1. Write down the probabilities mentioned above as formulas

Solution: Let m denote that it is morning and o denote that the TA is in the office.

1. $P(o \mid m) = \frac{1}{5}$
2. $P(m \mid o) = \frac{1}{3}$
3. $P(m) = P(\neg m) = \frac{1}{2}$

---

2. Compute the full joint probability distribution

Solution:

- $P(o, m) = P(o \mid m) \cdot P(m) = \frac{1}{10}$ *(product rule)*
- $P(\neg o, m) = \frac{4}{10}$, because $P(m) = P(o, m) + P(\neg o, m)$ *(marginalization)*

Now, from $P(m \mid o) \cdot P(o) = P(m, o)$ it follows that $P(o) = \frac{\frac{1}{10}}{\frac{1}{3}} = \frac{3}{10}$. So we get

- $P(o, \neg m) = \frac{2}{10}$, because $P(o) = P(o, m) + P(o, \neg m)$ *(marginalization)*
- $P(\neg o, \neg m) = \frac{3}{10}$, because $1 - P(o) = P(\neg o) = P(\neg o, m) + P(\neg o, \neg m)$ *(marginalization)*

---

3. What's the probability you'll meet your TA, if you come to the office in the afternoon?

Solution: $P(o \mid \neg m) = \frac{P(o,\neg m)}{P(\neg m)} = \frac{4}{10}$

Problem 2.2 (Stochastic and Conditional independence)

Consider the following random variables:

- three flips C_1, C_2, and C_3 of the same fair coin, which can be heads or tails
- the variable E which is 1 if both C_1 and C_2 are heads and 0 otherwise
- the variable F which is 1 if both C_2 and C_3 are heads and 0 otherwise

Out of the above 5 random variables,

1. Give three random variables X, Y, Z such that X and Y are stochastically independent but not conditionally independent given Z,

---

Solution: E.g., C_1 and C_2 with $Z = E$.

---

2. Give three random variables X, Y, Z such that X and Y are not stochastically independent but conditionally independent given Z.

Solution: E.g., E and F with $Z = C_2$.

Problem 2.3 (Calculations)

Assume *random variables* X, Y both with *domain* $\{0, 1, 2\}$, whose *joint probability distribution* $P(X, Y)$ is given by

x	y	$P(X = x, Y = y)$
0	0	a
0	1	b
0	2	c
1	0	d
1	1	e
1	2	f
2	0	g
2	1	h
2	2	i

1. Give all *subsets* of the *probabilities* $\{a, b, c, d, e, f, g, h, i\}$ that *sum* to 1.

   Solution: Only $\{a, b, c, d, e, f, g, h, i\}$

   ---

2. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X \neq 0)$.

   Solution: $d + e + f + g + h + i$

   ---

3. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 1, Y = 0)$.

   Solution: d

   ---

4. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 1 \mid Y = 0)$.

   Solution: $d/(a + d + g)$

   ---

5. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2)$.

   Solution: $c + e + g$

   ---

6. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2 \mid X > Y)$.

   Solution: $g/(d + g + h)$

Problem 2.4 (AFT Tests)

Trisomy 21 (*Down syndrome*) is a genetic anomaly that can be diagnosed during pregnancy using an amniotic fluid test.

The probability of a foetus having Down syndrome is strongly correlated with the age of the pregnant parent. We will only consider the following two age groups.

1. For 25 year olds the probability is one in 1250,
2. for 43 year old parents it increases to one in fifty.

However, diagnostic tests are never perfect. We distinguish two kinds of errors:

3. Type I Error (False Positive): The test result is positive even though the child is healthy.

4. Type II Error (False Negative): The test result is negative even though the child has trisomy 21.

The probabilities of Type I and Type II Errors are both merely 1% for amniotic fluid tests for Down syndrome.

1. Express the four items above in the form of conditional probabilities. Use the random variable F with domain $\{Age_{25}, Age_{43}\}$ for the age of the pregnant person and the Boolean random variables Pos and $Down$ for the propositions "*The amniotic fluid test is positive*" and "*The child has Down syndrome*" respectively.

Solution: $P(Down \mid F = Age_{25}) = 0.0008$, $P(Down \mid F = Age_{43}) = 0.02$, $P(Pos \mid \neg Down) = 0.01$, $P(\neg Pos \mid Down) = 0.01$.

---

2. Assume that we have a 25 year old pregnant person. Using Bayes' theorem, express and compute the probability that their child has Down syndrome, given that the amniotic fluid test is positive. What can we conclude from the result?

Solution: We normalize to $F = Age_{25}$, making $P(Down) = 0.0008$ and compute:

$$
\begin{aligned}
P(Down \mid Pos) &= \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos)} = \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos \wedge Down) + P(Pos \wedge \neg Down)} \\
&= \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos \mid Down) \cdot P(Down) + P(Pos \mid \neg Down) \cdot P(\neg Down)} \\
&= \frac{(1 - P(\neg Pos \mid Down)) \cdot P(Down)}{(1 - P(\neg Pos \mid Down) \cdot P(Down)) + P(Pos \mid \neg Down) \cdot (1 - P(Down))} \\
&= \frac{0.99 \cdot 0.0008}{0.99 \cdot 0.0008 + 0.010.9992} \approx 0.07
\end{aligned}
$$

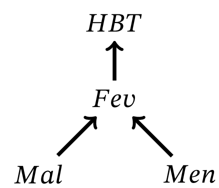
So, even with a positive test result, the probability of the child actually having Down syndrome is still only 7%, simply due to Down syndrome being relatively rare in young parents. Consequently, there is little point in applying this particular test without exceptional cause for concern.

Problem 2.5 (Medical Bayesian Network)

Both Malaria and Meningitis can cause a fever, which can be measured by checking for a high body temperature. Of course you may also have a high body temperature for other reasons. We consider the following random variables for a given patient:

- Mal: The patient has malaria.
- Men: The patient has meningitis.
- HBT: The patient has a high body temperature.
- Fev: The patient has a fever.

Consider the following Bayesian network for this situation:



1. Explain the purpose of the edges in the network regarding the conditional probability table.

Solution: The parents (i.e, nodes from which there are incoming edges) of X are the variables that X may depend on. The conditional probability table for X must take all of those as additional inputs.

---

2. What would have happened if we had constructed the network using the variable order Mal, Men, HBT, Fev? Would that have l better network?

Solution: We would have obtained additional edges from Mal and Men to Fev because they affect the probability of fever. That would be a worse network because more edges increase the complexity.

---

3. How do we compute the probability distribution for the patient having malaria, given that he has high body temperature? State the query variables, hidden variables and evidence and write down the equation for the probability we are interested in.

Solution: Query variable: Mal. Evidence: HBT. Hidden variables: Men, Fev. We get:

- start

$$P(Mal \mid HBT = true)$$

- normalization to turn the conditional distribution into an unconditional one

$$= \alpha P(Mal, HBT = true)$$

  where $\alpha = 1/P(HBT = true)$ is the constant factor that normalizes the vector $\langle P(Mal = true, HBT = true), P(Mal = false, HBT = true)\rangle$
- marginalization to bring in the hidden variables

$$= \alpha \sum_{m,f} P(Mal, HBT = true, Men = m, Fev = f)$$

  where m and f range over the possible values of Men and Fev
- chain rule to turn the joint distribution into a product of conditional ones, ordering the variables according to the structure of the network

$$= \alpha \cdot \sum_{m,f} P(Mal) \cdot P(Men = m \mid Mal) \cdot P(Fev = f \mid Mal, Men = m) \cdot$$

$$P(HBT = true \mid Fev = f, Mal, Men = m)$$

  Note that each factor is a vector with two entries, one for $Mal = true$ and one for $Mal = false$. These vectors are multiplied componentwise.
- use the structure of the network to drop redundant conditions

$$= \alpha \cdot \sum_{m,f} P(Mal) \cdot P(Men = m) \cdot P(Fev = f \mid Mal, Men = m) \cdot P(HBT = true \mid Fev = f)$$

  Now all factors are entries of the conditional probability table of the network, which can be plugged in to compute the result.

Because Mal is boolean, we could have started with $P(Mal = true \mid HBT = true)$ right away. Then we could have skipped the normalization step and would not have to multiply vectors. But for variables with many values, the above is practical because it derives the entire distribution in one go.

Problem 2.6 (Bayesian Networks in Python)

The goal of this exercise is to *implement* inference by enumeration in Bayesian networks in Python. You can find the necessary files at `https://kwarc.info/teaching/AI/resources/AI2/bayes/`.

Your task is to *implement* the `query` function in `bayes.py`. Use `test.py` for testing your *implementation*.

Important: We will test your code automatically. So please make sure that:

- The tests in `test.py` work on your code (without any modifications to `test.py`)
- You use a recent Python version ($\geq$ 3.5)
- You don't use any libraries

- You only upload a single file `bayes.py` with your *implementation* of `query`

Otherwise you risk getting no points.

Hint: First *implement* a function for the *full joint probability distribution*.

Problem 2.7 (Bayesian Epistemology)

Consider the following sayings. How can we express them in the form of conditional probabilities? Give actual *mathematical* formulas and explain how they relate to the sayings.

Are they (as statements about probabilities) actually true or under which assumptions can they be?

1. The simplest explanation is always the best.

Solution: Obviously, we have a problem to quantify what *simple* exactly means. So let's take one step back and think about what it means for one hypothesis A to be simpler than some hypothesis B.
One way to answer that question would be to say: A is *simpler* than B if the set of propositions entailed by A is a proper subset of those entailed by B. In this case we can say $AP_1 \wedge ... \wedge P_n$ and $BP_1 \wedge ... \wedge P_n \wedge P_{n+1} \wedge ... \wedge P_m$.
Naturally, we have $P(A) \geq P(A \wedge B)$ for any propositions A, B, hence under this interpretation the claim is true.

---

2. Extraordinary claims require extraordinary evidence.

Solution: Let's assume "extraordinary" means that the prior probability (in the absence of any evidence) is rather small, i.e. $P(A) \approx 0$[1]. If we want the claim A to be *likely*, we need to find some evidence e such that $P(A \mid e) \approx 1$. By Bayes' Theorem:

$$1 \approx P(A \mid e) = \frac{P(e \mid A) \cdot \overbrace{P(A)}^{\approx 0}}{P(e)}$$

It is immediately obvious that for this to hold, $P(e)$ needs to be highly unlikely, i.e. extraordinary.

---

3. Absence of evidence is not evidence of absence.

[1] I'll write $P(x) \approx 0$ resp. ≈ 1 simply for "is very unlikely" and "is very likely"

Solution: Let's assume that e is evidence *for* A iff $P(A \mid e) > P(A)$, i.e. observing e actually makes the proposition A more likely. I claim: *If e is evidence for A, then $\neg e$ is evidence for $\neg A$*, making the claim false:

$$
\begin{aligned}
& P(A \mid e) = \frac{P(e \mid A)P(A)}{P(e)} > P(A) \\
\rightsquigarrow & P(e \mid A)P(A) > P(e)P(A) \\
\rightsquigarrow & (1 - P(\neg e \mid A))P(A) > (1 - P(\neg e))P(A) \\
\rightsquigarrow & \underbrace{P(\neg e \mid A)P(A)}_{=P(\neg e, A)} < P(\neg e)P(A) \\
\rightsquigarrow & \overbrace{P(A \mid \neg e)\color{red}{P(\neg e)}} < \color{red}{P(\neg e)}P(A) \\
\rightsquigarrow & P(A \mid \neg e) < P(A) \\
\rightsquigarrow & 1 - P(\neg A \mid \neg e) < 1 - P(\neg A) \\
\rightsquigarrow & P(\neg A \mid \neg e) > P(\neg A)
\end{aligned}
$$

□