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

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

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.
2. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X \neq 0)$.
3. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 1, Y = 0)$.
4. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 1 \mid Y = 0)$.
5. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2)$.
6. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2 \mid X > Y)$.

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.
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?

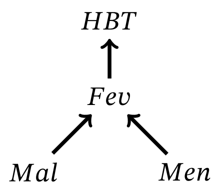
Problem 2.5 (Medical Bayesian Network)

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.
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?
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.

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 (≥ 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.
2. Extraordinary claims require extraordinary evidence.
3. Absence of evidence is not evidence of absence.