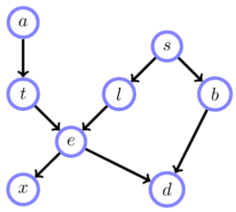
Exercise Session 2

1 Fred's Lisp Dilemma

Fred is debugging a LISP program. He just typed an expression to the LISP interpreter and now it will not respond to any further typing. As far as Fred knows, there are only two situations that could cause the LISP interpreter to stop running: (1) there are problems with the computer hardware; (2) there is a bug in Fred's code. Fred is also running an editor in which he is writing and editing his LISP code; if the hardware is functioning properly, then the text editor should still be running with a probability of 95%. When the hardware isn't functioning properly, the text editor could still be running with a probability of 10%. Additional information is that the hardware is pretty reliable, and is OK about 99% of the time, whereas Fred's LISP code is often buggy, say 40% of the time. In the case of bad hardware and buggy code, the interpreter stops running with a probability of 99%. In the case of good hardware and buggy code, the interpreter stops running with a probability of 40 %. In the case of good code and bad hardware, the interpreter stops running with a probability of 95%. In the case of good code and good hardware, the interpreter stops running with a probability of 10%.

1. Modeling Bayesian Networks:
    - Model the example above.
    - Write down the factorization for the joint probability distribution.
    - Calculate the number of values that need to be stored when the full joint is used.
    - Calculate the number of values that need to be stored when the factorized probability distribution is used.
2. Calculating probabilities in Bayesian Networks:
    - Calculate the probability that there is a bug in Fred's code when we know that the interpreter stopped working.
3. Independence in Bayesian Networks:
    - List all the pairs of variables which are independent.
    - List all the pairs of variables which are conditionally independent given exactly one other variable.



$x =$ Positive X-ray
$d =$ Dyspnea (Shortness of breath)
$e =$ Either Tuberculosis or Lung Cancer
$t =$ Tuberculosis
$l =$ Lung Cancer
$b =$ Bronchitis
$a =$ Visited Asia
$s =$ Smoker

$p(a = \mathsf{tr})$	$= 0.01$	$p(s = \mathsf{tr})$	$= 0.5$
$p(t = \mathsf{tr} \mid a = \mathsf{tr})$	$= 0.05$	$p(t = \mathsf{tr} \mid a = \mathsf{fa})$	$= 0.01$
$p(l = \mathsf{tr} \mid s = \mathsf{tr})$	$= 0.1$	$p(l = \mathsf{tr} \mid s = \mathsf{fa})$	$= 0.01$
$p(b = \mathsf{tr} \mid s = \mathsf{tr})$	$= 0.6$	$p(b = \mathsf{tr} \mid s = \mathsf{fa})$	$= 0.3$
$p(x = \mathsf{tr} \mid e = \mathsf{tr})$	$= 0.98$	$p(x = \mathsf{tr} \mid e = \mathsf{fa})$	$= 0.05$
$p(d = \mathsf{tr} \mid e = \mathsf{tr}, b = \mathsf{tr})$	$= 0.9$	$p(d = \mathsf{tr} \mid e = \mathsf{tr}, b = \mathsf{fa})$	$= 0.7$
$p(d = \mathsf{tr} \mid e = \mathsf{fa}, b = \mathsf{tr})$	$= 0.8$	$p(d = \mathsf{tr} \mid e = \mathsf{fa}, b = \mathsf{fa})$	$= 0.1$

$p(e = \mathsf{tr} \mid t, l) = 0$ *only if both t and l are* fa*, 1 otherwise.*

Figure 1: Bayesian network modeling (in)dependences of pulmonary diseases.

2 Independence in Bayesian Network

Consider the graphical probabilistic model in Figure 1 and answer questions:

1. Are the following independence statements true or false?
   - $t \perp\!\!\!\perp s \mid d$
   - $l \perp\!\!\!\perp b \mid s$
   - $a \perp\!\!\!\perp s \mid l$
   - $a \perp\!\!\!\perp s \mid l, d$
2. Calculate by hand the values for:
   - $p(d)$
   - $p(d \mid s = true)$
   - $p(d \mid s = false)$

3 Independence of Random Variables

There is a synergistic relationship between asbestos (A) exposure, smoking (S) and cancer (C). A model describing this relationship is given by $p(A, S, C) = p(C \mid A, S)p(A)p(S)$

- Is $A \perp\!\!\!\perp S \mid \emptyset$?
- Is $A \perp\!\!\!\perp S \mid C$?
- How could you adjust the model to account for the fact that people who work in the building industry have a higher likelihood to also be smokers and also a higher likelihood to asbestos exposure?
- Draw the Bayesian network representing the adjusted model.