CS267A: Probabilistic Programming and Relational Learning	Spring 2020

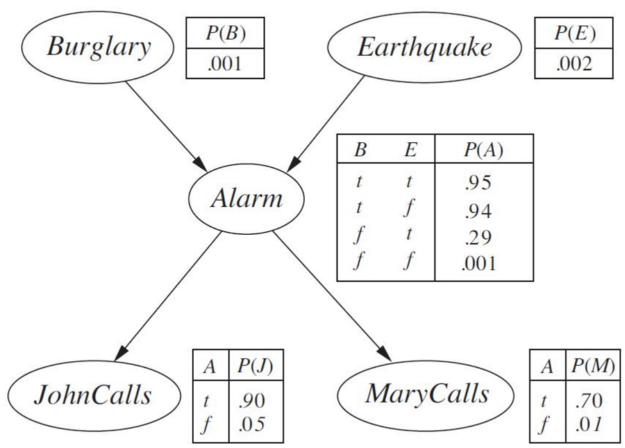
Homework 5: Due Monday May 18, 11:59PM

Instructions: Upload one file to CCLE: a PDF typeset using LaTeX containing your solutions (**including code**). Consider using package *listings* to embed the code. No late submissions will be accepted. See the syllabus for policies about collaboration and academic honesty.

Problem 1

Topic: Learning in ProbLog

The classic "earthquake, burglary" example is back. Please use ProbLog to model it.



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Consider we are the police station, which means, we are able to collect data regarding whether there is burglary/earthquake and whether Mary/John calls, but we don't know whether the alarms in their home sound or not. Please update the conditional probabilities (i.e. parameters) in your ProgLog program through Bayesian learning after observing the following examples.

- No Burglary, No earthquake, Mary did not call, John did not call
- Burglary, No earthquake, John called, Mary did not call
- No Burglary, No earthquake, Mary did not call, John did not call
- No Burglary, no earthquake, John did not call, Mary called
- Burglary, No earthquake, John called, Mary called
- Burglary, earthquake, John called, Mary called
- No Burglary, No earthquake, Mary did not call, John did not call
- No Burglary, No earthquake, Mary called, John did not call

Answer the following queries and provide your code along with your answer.

1. What is the probability that there is a burglary or earthquake given both John and Mary call?
2. What is the probability that there is an earthquake given no burglary and Mary calls?

Problem 2

Consider the following collection of deterministic databases with a single table called T having a single column, and a domain of A, B, and C:

$$\omega_1 = \begin{array}{|c|}\hline \text{A} \\ \hline \text{B} \\ \hline \text{C} \\ \hline\end{array} \qquad \omega_2 = \begin{array}{|c|}\hline \text{A} \\ \hline \text{B} \\ \hline\end{array} \qquad \omega_3 = \begin{array}{|c|}\hline \text{B} \\ \hline \text{C} \\ \hline\end{array} \qquad \omega_4 = \begin{array}{|c|}\hline \text{A} \\ \hline \text{C} \\ \hline\end{array}$$

$$\omega_5 = \begin{array}{|c|}\hline \text{A} \\ \hline\end{array} \qquad \omega_6 = \begin{array}{|c|}\hline \text{B} \\ \hline\end{array} \qquad \omega_7 = \begin{array}{|c|}\hline \text{C} \\ \hline\end{array} \qquad \omega_8 = \begin{array}{|c|}\hline \\ \hline\end{array}$$

(empty table)

Then we define a probabilistic database by assigning a probability to each of the above deterministic databases:

$$\Pr(\omega_1) = 0.16 \quad \Pr(\omega_2) = 0.16 \quad \Pr(\omega_3) = 0.24 \quad \Pr(\omega_4) = 0.04 \quad \Pr(\omega_5) = 0.04 \quad \Pr(\omega_6) = 0.24 \quad \Pr(\omega_7) = 0.06 \quad \Pr(\omega_8) = 0.06$$

1. What is $\Pr(\exists x.T(x))$ for the above probabilistic database?
2. Fill in the probabilities of the following tuple-independent database so that this tuple-independent database describes the same probabilistic database as above:

	$\Pr(\cdot)$
A	
B	
C	

Problem 3

Topics: Probabilistic Databases

Consider the following probabilistic database, which is similar to the one from class:

Let $\Pr(\omega_1) = 1/8$, $\Pr(\omega_2) = 1/4$, $\Pr(\omega_3) = 3/8$, and $\Pr(\omega_4) = 1/4$. Assume that the domain of the "Name" column is {Alice, Carol} and that the domain of the "Assoc." column is {Pixar, UPenn, Brown, INRIA}. Under the domain-closure assumption, answer the following:

Problem 3 continued on next page...

Name	Assoc.
Alice	Pixar
Carol	UPenn

(a) ω_1

Name	Assoc.
Alice	Pixar
Carol	INRIA

(b) ω_2

Name	Assoc.
Alice	Brown

(c) ω_3

Name	Assoc.

(d) ω_4

Figure 1: Probabilistic database with a single table called T.

1. Compute each of the following queries on the database:
   (a) $\Pr(T(\text{Alice}, \text{Pixar}))$.
   (b) $\Pr(\exists x.\ T(\text{Alice}, x))$.
   (c) $\Pr(\exists x, y.\ T(x, y))$.
   (d) $\Pr(\forall x.\ T(x, \text{Brown}))$.
2. Is this probabilistic database tuple-independent? Why or why not?

Problem 4

Topic: Probabilistic Database, ProbLog

Consider the following tuple-independent probabilistic database:

Student	Course	$\Pr(\cdot)$
Alice	Artificial Intelligence (AI)	0.7
Alice	Programming Languages (PL)	0.4
Bob	Artificial Intelligence (AI)	0.3

(a) The enrollment table, E.

Student	$\Pr(\cdot)$
Alice	0.7
Bob	0.4
Charlie	0.9

(b) The honors table, H.

Figure 2: A tuple-independent probabilistic database with two tables.

Part A

A tuple-independent probabilistic database describes a probability distribution on a collection of classical *deterministic* databases (the possible worlds). For just the table T_H, give (1) the set of all possible world databases of this table; (2) the probability of each such deterministic database.

Part B

Compute each of the following queries for the above probabilistic database *without* enumerating all possible worlds, and instead exploiting independence:

1. $\Pr(\exists x.H(x))$
2. $\Pr(\exists x.H(x) \wedge E(x, \text{Artificial Intelligence(AI)}))$
3. $\Pr(\exists x \exists y.H(x) \wedge E(x, y))$

Problem 4 continued on next page. . .

Part C

Implement the queries above in ProbLog, and make sure they give you the same numbers as before. Include your code.