Exam Probabilistic Programming 2019

April 18, 08:45–11:45

General Information:

- Mark every sheet with your **student number**.
- Check that your copy of the exam consists of **7 exercises**.
- This is an **open exam**, i.e. slides are permitted. However, solutions to the exercises and your own notes are not permitted.
- Write with blue or black ink; do **not** use a pencil or red ink.
- Make sure all electronic devices are switched off and are nowhere near you.
- Any attempt at deception leads to failure for this exam, even if detected only later.

Exercise 1 (Expected Runtimes) 8%

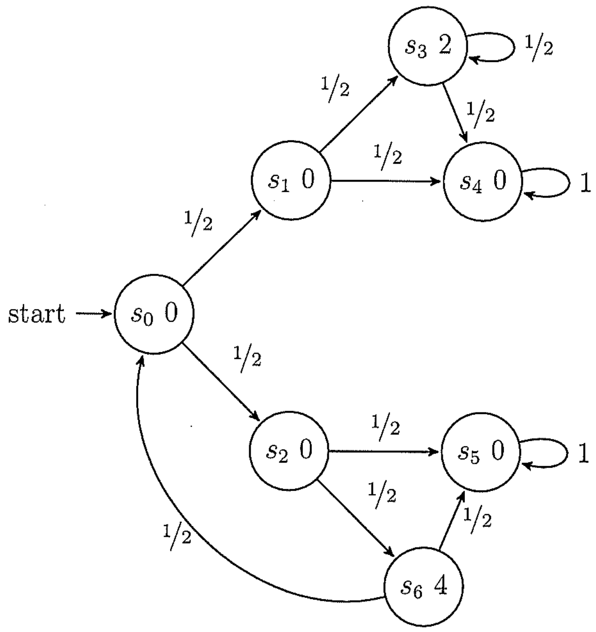
Apply the ert–calculus to compute the expected runtime of the program:

```
if (y ≠ 0) {
    {x := x · y} [1/2] {y := 3}
} else {
    {x := 1} [1/3] {x := 3}
};
if (x ≥ 1) {
    {y := 2} [1/2] {x := 3}
} else {
    skip;
    x := 3;
    skip;
    {y := x + 1} [1/4] {y := 2 · y};
    skip
};
{x := x^2} [1/2] {y := y^3}
```

Hint: you may insert your results in the program text above.

Exercise 2 (Conditional Expected Rewards) **12%**

Consider the following Markov chain M with rewards, where the reward of each state is provided next to the state's name (inside the circles).



(a) [4%] Compute $\mathrm{ER}^M(\lozenge\{s_4\})$.

(b) [4%] Compute $\mathrm{Pr}^M(\neg\lozenge\{s_5\})$.

(c) [4%] Compute $\mathrm{ER}^M(\lozenge\{s_4\} \mid \neg\lozenge\{s_5\})$.

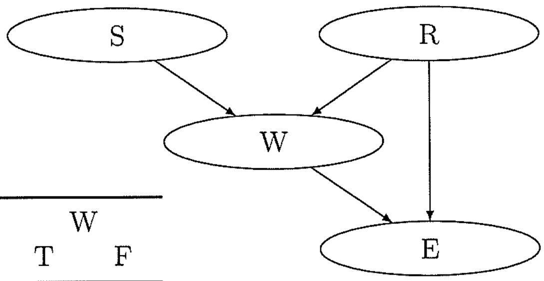
Provide intermediate steps such that your computations are comprehensible.

Hint: expected rewards can be obtained by solving a system of linear equations which is similar (but not identical) to computing reachability probabilities.

Exercise 3 (Sampling Time of Bayesian Networks) **12%**

Consider the Bayesian network B:

S	
T	F
0.25	0.75



R	
T	F
0.6	0.4

		W	
S	R	T	F
F	F	0.61	0.39
F	T	0.93	0.07
T	F	0.68	0.32
T	T	0.83	0.17

		E	
W	R	T	F
F	F	0.48	0.52
F	T	0.11	0.89
T	F	0.65	0.35
T	T	0.26	0.74

(a) [4%] Assume we observe that $E = T$, $W = F$ and $R = F$. Apply the transformation from the lecture to translate the Bayesian network B and the above observations into an equivalent pGCL program.

(b) [8%] Determine the expected sampling time of Bayesian network B for the observations $E = T$, $W = F$ and $R = F$.

Exercise 4 (Loops and Invariants) 16%

(a) [12%] Consider the following pGCL program P:

$$
\begin{array}{l}
\texttt{while}\,(x > 0)\,\{ \\
\quad y := [\texttt{true}\rangle \cdot 0.8 \,+\, [\texttt{false}\rangle \cdot 0.2; \\
\quad \texttt{if}\,(y)\,\{ \\
\quad\quad z := [17\rangle \cdot 0.5 \,+\, [23\rangle \cdot 0.5; \\
\quad\quad \texttt{while}\,(z \neq 0)\{ \\
\quad\quad\quad \{\texttt{skip}\}\,[0.3]\,\{z := 0\} \\
\quad\quad \}; \\
\quad\quad c := c + 1 \\
\quad \}\,\texttt{else}\,\{\ x := 0\ \} \\
\}
\end{array}
$$

Show that $I = c + 4 \cdot [x > 0]$ is an upper *wp*-invariant of P with respect to post-expectation c. That is, prove that $\Phi_c(I) \leq I$, where Φ_c is the *wp*-characteristic function of P with respect to post-expectation c.

(b) [4%] Consider the loop $P = \texttt{while}(G)\,\{P'\}$, where G is a guard and P' is a loop-free pGCL program. Moreover, let Φ_f be the *wp*-characteristic function of P with respect to postexpectation $f \in \mathbb{E}_{\leq 1}$.

Show that Φ_f has a unique fixed point (in $\mathbb{E}_{\leq 1}$) if and only if

$$wp(P, f) \;=\; wlp(P, f)\ .$$

Exercise 5 (Conditioning)

15%

(a) [5%] Consider the pGCL program P below:

```
1 :  x := [1⟩ · 0.3 + [2⟩ · 0.2 + [3⟩ · 0.5;
2 :  if(x < 3) {
3 :      y := [4⟩ · 0.7 + [0⟩ · 0.3
4 :  } else {
5 :      y := 0
6 :  };
7 :  observe(x + y > 3);
8 :  z := [0⟩ · 0.2 + [1⟩ · 0.8;
9 :  observe(x + y > 5)
```

Construct the Markov chain corresponding to P with initial state σ defined by $\sigma(x) = 1$, $\sigma(y) = 2$, and $\sigma(z) = 3$.

Hint: You may use line number ℓ to indicate the fragment starting at line ℓ.

(b) [10%] Apply the *cwp*-calculus to compute (the quotient corresponding to) $cwp(P, [x = 2])$.

Exercise 6 (Probabilitic databases) **25%**

Consider the example data of Figure 1. It concerns a database filled with the result of integrating two sources that each contain the information of one music album. Because of some typos (one or two t's in "Nighttime birds"; one or two s's in "Confusion"), it is uncertain whether they are the same album or not. Furthermore, if they would be the same album, then some data about the album is uncertain, because the two sources disagree with each other. The table **album** contains the integrated album(s) and the table **song** contains the songs of these album(s).

album

⟨**aid, title, artist**⟩	φ
⟨1, Nighttime birds, The Gathering⟩	$(s = 0 \wedge \neg w = 3) \vee (s = 1 \wedge w = 1)$
⟨1, Nightime birds, The Gathering⟩	$s = 1 \wedge w = 2$
⟨2, Nightime birds, The Gathering⟩	$s = 0 \wedge \neg w = 3$

song

⟨**sid, aid, name**⟩	φ
⟨1, 1, On Most Surfaces⟩	$\neg w = 3$
⟨2, 1, Confusion⟩	$w = 1$
⟨3, 1, The May Song⟩	$w = 1$
⟨2, 1, Confussion⟩	$w = 2$
⟨1, 2, On Most Surfaces⟩	$\neg w = 3$
⟨2, 2, Confussion⟩	$w = 2$

φ	p	**description**
$s = 0$	0.1	Different albums
$s = 1$	0.9	One album
$w = 1$	0.6	Source 1 is correct
$w = 2$	0.3	Source 2 is correct
$w = 3$	0.1	Neither source is correct

Figure 1: Sample probabilistic integration result between two sources containing data on one album.

(a) [3%] What are the contents of the possible world belonging to $s = 0 \wedge w = 3$, i.e., which tables exist in this possible world, and which records are contained in these tables?

(b) [3%] How many possible worlds does the probabilistic database of Figure 1 contain? Explain your answer by giving the complete calculation of the answer.

(c) [3%] Calculate the probability of $(s = 0 \wedge \neg w = 3) \vee (s = 1 \wedge w = 1)$.

(d) [3%]
Given the probabilistic SQL query Q

```
SELECT s.title, CONF()
FROM album a, song s
WHERE a.aid=s.aid
  AND s.name="Confusion"
GROUP BY s.title
```

and the probabilistic algebra expression E

$$\pi_{\text{s.title}}(\bowtie_{\text{a.aid=s.aid}} (\text{album}, \sigma_{\text{s.name="Confusion"}}(\text{song})))$$

Which of the following statements are true?

1. Q and E are the same (i.e., have the same result)
2. Q and E are different: an *aggregation* is missing in E
3. Q and E are different: a *projection* is missing in E
4. Q and E are different: a *selection* is missing in E
5. E contains a *join*
6. The result of Q is certain.

(Note that there can be multiple true statements.)

(e) [8%] What is the exact result of probabilistic algebra expression E above?

(f) [5%] In indeterministic duplicate detection, an M-graph is constructed from similarity match results of a duplicate detection tool which ran on tuples a, b, c, d. The tool determines the following similarities: $s(a-b) = 0.95$, $s(a-c) = 0.7$, $s(b-c) = 0.8$, $s(a-d) = 0.6$, $s(b-d) = 0.3$, $s(c-d) = 0.3$. We set the upper threshold to 0.9 and the lower threshold to 0.5. How many possible worlds does this produce when we enforce these thresholds? Explain your answer based on the M-graph.

*Remark:*probabilities of possible worlds are not requested, so there is no need to compute them.

Exercise 7 (Healthiness Properties) $\boxed{12\%}$

Let P be a pGCL program. Moreover, let $f, g \in \mathbb{E}$ and $h \in \mathbb{E}_{\leq 1}$.

Prove or disprove:

(a) [4%] $wp(P, f+g) \;=\; \max\{wp(P, f), wp(P, g)\}$.

(b) [4%] If f is unaffected by P, then $wp(P, f) \geq f$.

(c) [4%] If h is unaffected by P, then $wlp(P, h) \geq h$.

Hint: Recall that f is unaffected by P iff $\mathit{Vars}(f) \cap \mathit{Mod}(P) = \emptyset$.

Here, $\mathit{Vars}(f)$ is the set of all variables in expectation f. Formally,

$$x \in \mathit{Vars}(f) \quad \text{iff} \quad \exists s \exists u \exists v \colon f(s[x := u]) \neq f(s[x := v]) \ .$$

Moreover, $\mathit{Mod}(P)$ is the set of all variables modified by P. That is, $x \in \mathit{Mod}(P)$ iff P contains an assignment of the form $x := \ldots$.