Exam Probabilistic Programming 2022/23

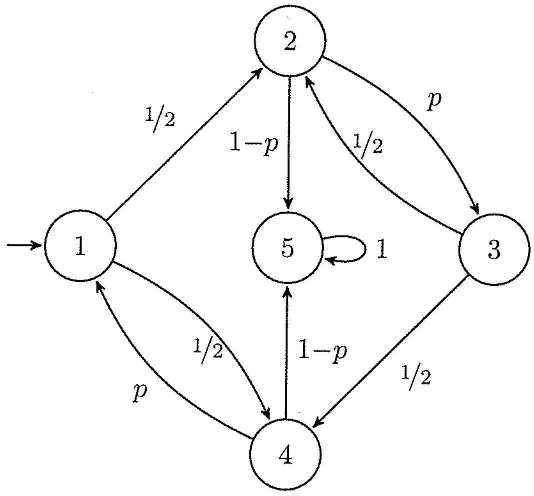
April 20, 2023, 08:45–11:45

General Information:

- Mark every sheet with your **student number**.
- Check that your copy of the exam consists of **five exercises**.
- This is an **open exam**, i.e. all lecture material (slides, exercises, solutions to the exercises) is permitted. This material is permitted in printed form only.
- No laptops, PDAs, mobile phones are allowed to use during the exam.
- Write with blue or black ink; do **not** use a pencil or red ink.
- You are neither allowed the help of anyone to complete your exam, nor is it allowed to help anyone else in completing this exam.
- Any attempt at deception leads to failure for this exam, even if detected later.

Exercise 1 (Markov Chains) $\boxed{18\%}$

(a) [6%] Consider the following Markov chain:



Let $p = \frac{1}{2}$ and $G = \{4\}$. Determine the set $\Sigma_?$ of states for which the probability to reach G is neither 0 nor 1. Then specify a matrix $\vec{M}$ of dimension $|\Sigma_?| \times |\Sigma_?|$ and a vector $\vec{b}$ such that the linear equation system $\vec{M}\vec{x} = \vec{b}$ has the unique solution:

$$\vec{x} = \big(\, Pr(1 \models \Diamond G)\,,\, Pr(2 \models \Diamond G)\,,\, Pr(3 \models \Diamond G) \,\big)\ .$$

Hint: You do not *have to compute $\vec{x}$, specifying $\Sigma_?$, $\vec{M}$ and $\vec{b}$ suffices.*

(b) [8%] Consider the following pGCL program P:

$$\{x := 0;\}\ [1/2]\ \{x := 1;\}$$
$$\{y := 0;\}\ [1/3]\ \{y := 1;\}$$
$$\mathsf{observe}\ (2x + y \leq 2)$$

Depict the reachable fragment of the Markov chain $[\![P]\!]$ (= operational semantics of P) from initial state $\langle P\rangle 00$. Here, "00" stands for the variable valuation where both x and y are assigned zero.

(c) [4%] Express the expected value of x after program termination in terms of a conditional expected reward in $[\![P]\!]$, and compute this expected reward.

Exercise 2 (Probabilistic Loops) $\boxed{22\%}$

Consider the following pGCL program P:

$$
\begin{array}{l}
y := 1; \\
\mathsf{while}\,(i > 0)\,\{ \\
\quad \{\mathsf{diverge}\}\ [1/2]\ \{y := y \cdot i\} \\
\quad i := i - 1 \\
\}
\end{array}
$$

Throughout this task we fix the post-expectation y. Moreover, we let Φ_y be the *wp* loop-characteristic functional of P with respect to post-expectation y.

(a) [6%] Let X be an arbitrary expectation. Give an expression for $\Phi_y(X)$.

Hint: Recall that $\mathsf{wp}[\![\mathit{diverge}]\!](X) = 0$.

Hint: Your solution may contain terms of the form $X[z/E]$ where z is a program variable and E is an arithmetic expression.

(b) [10%] Compute $\Phi_y^0(0)$, $\Phi_y^1(0)$, $\Phi_y^2(0)$, and $\Phi_y^3(0)$.

(c) [6%] Find an expression for the least fixed point of Φ_y.

Hint: You do not *have to prove that your answer is correct.*

Hint: Your expression may contain infinite sums, Iverson brackets, factorials (i.e., $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$), and basic arithmetic operations.

Exercise 3 (Program Equivalence) $\boxed{18\%}$

Consider the following two pGCL programs with constant parameters p and q ($p, q \in [0,1] \cap \mathbb{R}$):

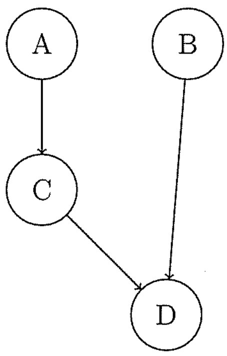
$y := 4$; $\mathsf{if}\ (y < 0)\ \{$ $\quad$ **skip** $\}\ \mathsf{else}\ \{$ $\quad \{x := 1\}\ [q]\ \{$**skip**$\}$ $\}$	$\{$ $\quad \{x := x + 3\}\ [p]\ \{x := 0\}$ $\}\ [q]\ \{$ $\quad x := 0$ $\}$
Program P_1	Program P_2

(a) [14%] Compute the weakest pre-expectations for programs P_1 and P_2 with respect to post-expectation x.

(b) [4%] Give all parameter values $p, q \in [0,1] \cap \mathbb{R}$ such that $\mathsf{wp}[\![P_1]\!](x) = \mathsf{wp}[\![P_2]\!](x)$, or prove that such values do not exist.

Exercise 4 (Bayesian Networks) $\boxed{17\%}$

Consider the Bayesian network given by the directed graph and conditional probability tables below.



$A=1$	$A=0$
0.9	0.1

$B=1$	$B=0$
0.5	0.5

	$C=1$	$C=0$
$A=1$	0.2	0.8
$A=0$	0.3	0.7

	$D=1$	$D=0$
$C=1, B=1$	0.4	0.6
$C=1, B=0$	0.5	0.5
$C=0, B=1$	0.6	0.4
$C=0, B=0$	0.7	0.3

Hint: In the following subtasks, it suffices to give numeric results in terms of basic arithmetic operations. You do not need to reduce fractions, for example.

(a) [3%] Compute the probability $P(A=1 \mid C=0)$.

(b) [6%] Write a pGCL program that simulates rejection sampling from the above Bayesian network with respect to the evidence $D=0$.

Hint: You may use a **repeat-until** *loop.*

(c) [8%] Compute the expected runtime of your loop from answer (b)).

Hint: You may use that $P(D=0) = \frac{49}{125}$ *holds for the Bayesian network.*

Hint: The ert-semantics of the **repeat-until** *loop are given by*

$$\mathsf{ert}[\![\mathbf{repeat}\,\{\,P\,\}\,\mathbf{until}\,(\varphi)]\!](f) = \mathsf{ert}[\![P;\ \mathsf{while}\,(\neg\varphi)\ \{P\}]\!](f)\ .$$

Hint: Use elementary properties of ert/wp. *They can save a lot of calculations.*

Exercise 5 (Probabilistic Databases) $\boxed{25\%}$

The example data of Figure 2 is about findings at a crime scene. There were four items found which are denoted with A, B, C, and D (the 'iid' attribute). Item 'A' is a gun, item 'B' is a coat, item 'C' is a shirt, and item 'D' is either a pair of slippers or a pair of flip flops. The officer reporting it is Dutch and they reported a pair of 'slippers' but we are unsure whether they reported in Dutch or in English: the Dutch word 'slippers' is 'flip flops' in English and the Dutch word 'pantoffels' is 'slippers' in English.

It is important for the investigation whether or not these items were found together as a group. The most reliable source says that they were all found together in the house (group 'ABCD'). But the Dutch officer claims she saw the slippers in the garden (i.e., two groups 'ABC' and 'D'). Some bystander (not really looking very reliable) who was with the Dutch officer says they saw the gun in the shed and not in the house (i.e., three groups 'A', 'BC', 'D').

This evidence was modeled using the *uncertain groupings* approach, which resulted in the example data of Figure 2. Groups are identified by the 'gid' attribute. The 'rel' table is just there to store which items are in which groups, which you can also derive from the letters in the 'gid'.

(a) [2%] How many possible worlds does the probabilistic database of Figure 2 contain? Explain your answer by giving the complete calculation of the answer.

(b) [3%] There are three groups that contain item A, the gun: groups 'A', 'ABC', and 'ABCD'. These groups are found in either the 'shed' or the 'house' according to the database. How can you see in the database that the gun was not both in the shed and the house at the same time?

(c) [3%] Calculate the probability of $(u = 2 \wedge y = 2) \vee x = 2$ (i.e., of the gun found in the shed or the slippers in the garden). Explain your answer by giving the complete calculation of the answer.

groups

$\langle$gid, location$\rangle$	φ
$\langle$A, shed$\rangle$	$u = 2 \wedge y = 2$
$\langle$ABC, house$\rangle$	$v = 2 \wedge y = 1 \wedge z = 1$
$\langle$ABCD, house$\rangle$	$u = 1 \wedge v = 1 \wedge w = 1 \wedge x = 1$
$\langle$BC, house$\rangle$	$w = 2 \wedge z = 2$
$\langle$D, garden$\rangle$	$x = 2$

items

$\langle$iid, item$\rangle$	φ
$\langle$A, gun$\rangle$	$\top$
$\langle$B, coat$\rangle$	$\top$
$\langle$C, shirt$\rangle$	$\top$
$\langle$D, slippers$\rangle$	$t = 1$
$\langle$D, flip flops$\rangle$	$t = 2$

rel

$\langle$gid, iid$\rangle$
$\langle$A, A$\rangle$
$\langle$ABC, A$\rangle$
$\langle$ABC, B$\rangle$
$\langle$ABC, C$\rangle$
$\langle$ABCD, A$\rangle$
$\langle$ABCD, B$\rangle$
$\langle$ABCD, C$\rangle$
$\langle$ABCD, D$\rangle$
$\langle$BC, B$\rangle$
$\langle$BC, C$\rangle$
$\langle$D, D$\rangle$

dictionary

φ	prob
$t = 1$	1/2
$t = 2$	1/2
$u = 1$	5/6
$u = 2$	1/6
$v = 1$	5/8
$v = 2$	3/8
$w = 1$	5/6
$w = 2$	1/6
$x = 1$	5/9
$x = 2$	4/9
$y = 1$	3/4
$y = 2$	1/4
$z = 1$	3/4
$z = 2$	1/4

Figure 2: Example probabilistic data on findings at a crime scene.

(d) [3%] (**hard question**) Suspicion arose whether or not this *uncertain groupings* approach was suitable for this situation. For this purpose, we are going to test a fact that we know to be true: the coat was found in the house. It turns out that the suspicion is justified: there are possible worlds in this database in which the coat is not in the house (it is in fact not in any location in such a world).
Give the sentence of exactly one possible world in which the coat is not in the house (i.e., a counter example). Explain your answer.
NB: If you cannot answer, then at least provide some thoughts; you might score a few points.

(e) [3%] Given the following DuBio query (the attribute containing the sentences is denoted with φ in the query)

```
SELECT i.item, g.location, i.φ & g.φ AS φ
FROM items i, groups g, rel r
WHERE i.iid=r.iid AND g.gid=r.gid
GROUP BY i.item, g.location
```

For each statement, indicate whether or not it is true *and provide an explanation why.*

- ○ The query produces an error because we are joining with table 'rel', but we did not include '`& r.`φ'.
- ○ The query produces an error because we use a '`GROUP BY`' without an attribute from the 'rel' table.
- ○ The query produces an error because we have a '`GROUP BY`', but we do not aggregate the sentences φ.
- ○ The dictionary, which keeps an administration of the random variables, their alternatives and probabilities, is not used in this query.

(f) [6%] Given the probabilistic algebra expression E below ('g.a' refers to attribute a of table "groups"; 'i.a' refers to attribute a of table "items"; 'r.a' refers to attribute a of table "rel"). Give the result of E.

$$E \equiv \pi_{\text{i.item,g.location}}(\bowtie_{\text{g.gid=r.gid}} (\text{groups}, \bowtie_{\text{i.iid=r.iid}} (\sigma_{\text{i.item='slippers'}}(\text{items}), \text{rel})))$$

NB: I do not ask for a derivation, only the result. Note that an important part of your answer is to have it in *the right form*, so take care to provide all components that a result of a probabilistic algebra expression should have and omit those components that such a result should not have.

(g) [5%] In indeterministic duplicate detection, an M-graph is constructed from similarity match results of a duplicate detection tool which ran on tuples a="Slippers", b="Slipper", c="Sniper", and d="Pantoffel". The tool determines the following similarities: $s(a-b) = 0.8$, $s(a-c) = 0.5$, $s(a-d) = 0.95$, $s(b-c) = 0.6$, $s(b-d) = 0.7$, $s(c-d) = 0.1$. We set the upper threshold to 0.9 and the lower threshold to 0.3.

[1]'Conjunction' simply means an 'and' of multiple things

1. Draw the M-graph *after the thresholds have been applied.*
2. Which possible worlds does this produce? Use the following notation: $\{\ldots\}$ for the set of records comprising a possible world; ab for the merge of records a and b (other combinations analogously). Explain your answer.

NB: I do not ask for probabilities of possible worlds, so no need to compute them.