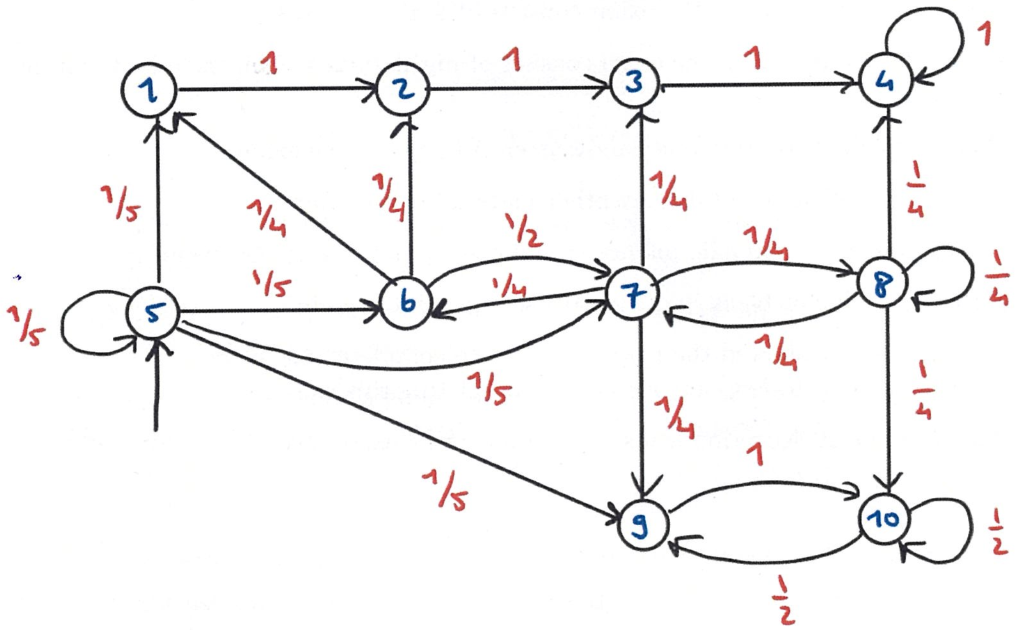
Exam April 17th 2025, 8:45-11:45

General remarks:

- Mark every sheet with your **student id number**.
- Check that your copy of the exam consists of **six exercises**.
- Check that your copy of the exam consists of **eight pages**. This includes the front page (this page).
- You are allowed to bring one handwritten A4 page to the exam.
- You are not allowed to take any other material to the exam.
- 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.

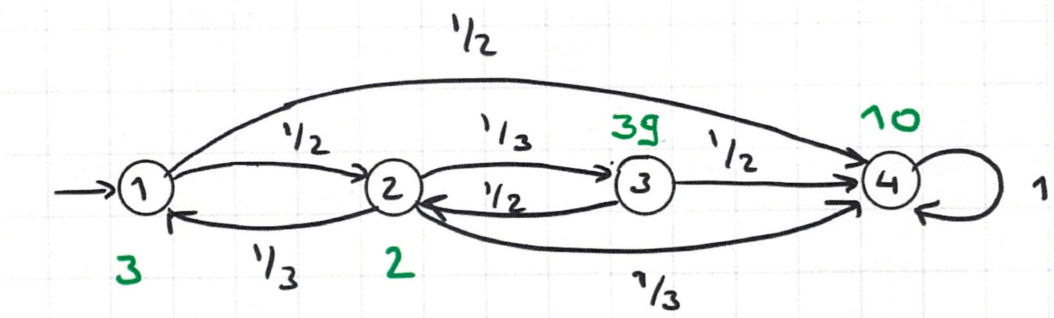
Exercise 1 (Reachability Probabilities and Rewards) 16%

Consider the Markov chain depicted below:



(a) [8%] Determine $\Pr(\Diamond\{8\})$ for all states in the Markov chain.

(b) [8%] Consider the Markov reward chain where rewards are indicated in green:

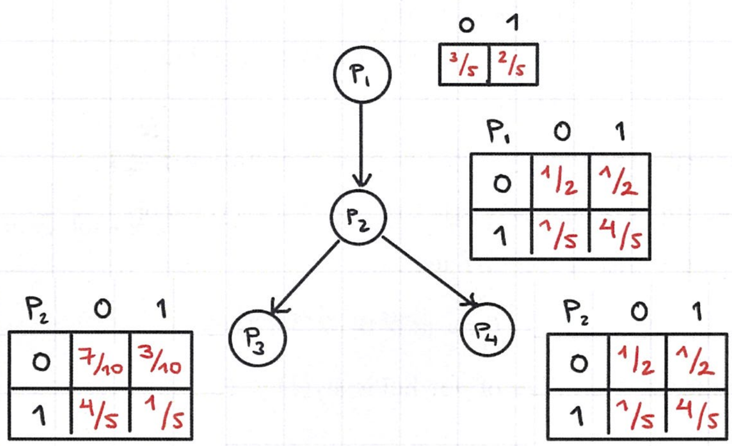


Determine $ER(\Diamond\{4\})$.

Provide intermediate steps such that your computations are comprehensible.

Exercise 2 **(Bayesian Networks)** 15%

Consider the following Bayesian network on four binary random variables:



(a) [4%] Determine $\Pr(p_1 \mid p_2 \wedge \neg p_3)$.

(b) [5%] Give a pGCL program for the Bayesian network with evidence $p_2 \wedge \neg p_3$.

(c) [6%] Determine the expected run-time of your pGCL program using:

$$ert(\texttt{repeat } Seq \texttt{ until } G, f) = \frac{1 + ert(Seq, [G].f)}{wp(Seq, [G])}$$

where *Seq* stands for the sequential composition of the programs for the Bayesian network vertices.

It suffices to indicate the expected run-times of the program fragments for the nodes in the Bayesian network without the detailed *ert* calculations.

Exercise 3 (A Random Walk) 16%

Consider the following pGCL program P:

$$
\begin{array}{l}
\texttt{while}\,(x > 0)\,\{ \\
\qquad \{x := x - 1\}\,[p]\,\{x := x + 1\} \\
\}
\end{array}
$$

where p is a probability in $\mathbb{Q} \cap [0, 1]$. Let post-expectation $f = \mathbf{1}$.

(a) [6%] Determine the characteristic function Φ_f for the loop.

(b) [10%] Let expectation I be defined as:

$$I \;=\; [x > 0] \cdot 2^{-x} + [x \leq 0]$$

Determine for which values of p it holds $\Phi_f(I) \leq I$.

Exercise 4 (Loops, Invariants, and Termination) $\boxed{18\%}$

Consider the following pGCL program P:

$$
\begin{aligned}
&x := 100;\\
&\texttt{while}\,(x = 100)\,\{\\
&\quad \{x := 10\}[1/10]\{y := y + \sqrt{10}\}\\
&\}
\end{aligned}
$$

Let expectation I_n for natural number n be defined by:

$$I_n \;:=\; [x \neq 100] + [x = 100] \cdot \sum_{k=0}^{n-1} 1/10 \cdot (9/10)^k \;,$$

where $\sum_{k=0}^{-1} \ldots = 0$.

Let post-expectation $f = \mathbf{1}$.

(a) [6%] Give the characteristic function Φ_f of the loop in P

(b) [4%] Show that $I_0 \;=\; \Phi_f(\mathbf{0})$

(c) [8%] Show that for all $n \geq 0$, it holds: $I_{n+1} \;=\; \Phi_f(I_n)$

Exercise 5 (The Meaning of Formal Specifications) 10%

Let F be a predicate. Match each of the following five statements about probabilistic program P with their corresponding colloquial description below. Note that there are more descriptions than statements. To each statement at most one description is correct.

Statements:

(a) $[F] \; = \; \mathsf{wp}[\![P]\!](\mathbf{1})$

(b) $[F] \; \sqsubseteq \; \mathsf{wp}[\![P]\!](\mathbf{1})$

(c) $\mathbf{1} \; = \; \mathsf{wlp}[\![P]\!](\mathbf{0})$

(d) $\mathbf{0} \; \sqsubseteq \; \mathsf{wp}[\![P]\!]([F])$

(e) $\mathsf{wp}[\![P]\!]([F]) \; = \; [\mathsf{wp}[\![P]\!](F)]$

Colloquial descriptions:

1. The statement is logically equivalent to true.
2. The statement is logically equivalent to false.
3. Program P diverges almost surely for all initial states.
4. Program P never terminates in a state satisfying F.
5. Program P almost surely terminates when started in a state satisfying F.
6. Program P almost surely terminates when started in a state satisfying F and diverges almost surely when started in a state that satisfies $\neg F$.
7. None of the descriptions above.

Exercise 6 (Probabilistic databases) 25%

The example data of Figure 1 contains voice recognition results from people pronouncing their name and a money amount in euros.

There are 2 tables: persons with a person identifier (pid), an amount identifier for the amount they pronounced (aid), and the recognized name (name). There is uncertainty about the actual names and their spelling that were pronounced. As you can imagine, it can be very hard for voice recognition software to distinguish between names like "Claire" ($p_1 = 1$) and "Clair" ($p_1 = 2$). Note that the system works in such a way, that if one person wants to say two amounts, they pronounce their name only once and then pronounce the two amounts.

The other table is amounts with an amount identifier for a pronounced amount (aid), the recognized amount as a number (amount), and the recognized amount as the word that was recognized (wordamount). Analogously, also around the recognition of amounts of money, the software may have a hard time to distinguish certain amounts especially amounts like 30 ($a_1 = 1$) and 13 ($a_1 = 2$) since they sound very much alike: "Thirty" vs. "Thirteen".

(a) [3%] How many possible worlds does the probabilistic database of Figure 1 represent?

(b) [3%] Calculate the probability of $(p_1 = 2 \vee p_4 = 1) \wedge a_1 = 1$.

Explain your answer by giving the complete calculation of the answer.

(c) [4%] Given the following DuBio query for the example database (the attribute containing the sentences is denoted with φ in the query)

amounts

⟨aid, amount, wordamount⟩	φ
⟨10, 30.00, "Thirty"⟩	$a_1 = 1$
⟨10, 13.00, "Thirteen"⟩	$a_1 = 2$
⟨11, 24.00, "Twenty-four"⟩	$\top$
⟨12, 60.00, "Sixty"⟩	$a_2 = 1$
⟨12, 16.00, "Sixteen"⟩	$a_2 = 2$

persons

⟨pid, aid, name⟩	φ
⟨1, 10, Claire⟩	$p_1 = 1$
⟨1, 10, Clair⟩	$p_1 = 2$
⟨2, 11, Catherine⟩	$p_2 = 1$
⟨2, 11, Kathryn⟩	$p_2 = 2$
⟨2, 11, Katharine⟩	$p_2 = 3$
⟨3, 12, Sean⟩	$p_3 = 1$
⟨3, 12, Shawn⟩	$p_3 = 2$
⟨4, 10, Sean⟩	$p_4 = 1$
⟨4, 10, Shawn⟩	$p_4 = 2$

dictionary

φ	prob
$a_1 = 1$	1/2
$a_1 = 2$	1/2
$a_2 = 1$	1/6
$a_2 = 2$	5/6
$p_1 = 1$	1/2
$p_1 = 2$	1/2
$p_2 = 1$	1/5
$p_2 = 2$	2/5
$p_2 = 3$	2/5
$p_3 = 1$	1/3
$p_3 = 2$	2/3
$p_4 = 1$	3/4
$p_4 = 2$	1/4

Figure 1: Example probabilistic data on voice recognition results of people pronouncing their name and a money amount in euros.

```
SELECT p.name, a.amount, p.φ & a.φ AS _sentence
FROM persons p, amounts a
WHERE p.aid=a.aid AND a.amount≥ 30
```

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

- ○ The result of this query is probabilistic data.
- ○ The reason for the '&' is the SELECT-clause is because the query computes a join over two tables.
- ○ The query produces an error because we do not have a 'GROUP BY', which is always needed in DuBio queries.
- ○ The query constructs a sentence for each possible answer; the sum of the probabilities of all these sentences is 1.

(d) [6%] Given the probabilistic algebra expression E below ('p.a' refers to attribute a of table "persons"; 's.a' refers to attribute a of table "amounts"). The δ operation is for duplicate removal
Give the result of E.

$$E \equiv \delta(\pi_{\text{a.wordamount,a.amount}}(\bowtie_{\text{p.aid=a.aid}} (\text{persons}, \sigma_{\text{a.amount}<20}(\text{amounts}))))$$

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.

(e) [4%] As you can see in the example data, two times a person Sean/Shawn pronounced a euro amount. Given the records and sentences in the example data, was it the same person Sean/Shawn or two different persons who happened to have the same name Sean/Shawn? Explain your answer.

Furthermore, how would you change the records and/or sentences to represent the other situation? Explain your answer. In your explanations, be as concrete and explicit as possible.

(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="Catherine", b="Kathryn", c="Katharine", d="Catarina". The tool determines the following similarities: $s(a-b) = 0.6$, $s(a-c) = 0.8$, $s(a-d) = 0.2$, $s(b-c) = 0.6$, $s(b-d) = 0.2$, $s(c-d) = 0.5$. We set the upper threshold to 0.9 and the lower threshold to 0.4.

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.