Exercise Sheet 3

General remarks:

- **Due date:** November 11th 12:30 (before the exercise class).
- Please submit your solutions via Moodle. Remember to provide your matriculation number. It is necessary to hand in your solutions in groups of **three**. You may use the Moodle forum to form groups.
- Solutions must be written in English.
- While we will publish sketches of exercise solutions, we do *not* guarantee that these sketches contain all details that are necessary to properly solve an exercise. Hence, it is recommended to attend the exercise classes.
- If you have any questions regarding the lecture or the exercise, please use the forum in Moodle.

Exercise 1 (Metropolis-Hastings) **25P**

(a) [10P] Write a WebPPL program that given a function f and a *symmetric* sampling kernel $g(x_{t+1}|x_t)$, implements the Metropolis-Hastings algorithm.

Solution:

```
var sampler = function(n, func, prop, x_old, samples){
  if(n == 0) return samples;
  var x_new = prop(x_old);
  var alpha = func(x_new) / func(x_old) // symmetric kernel
  var u = uniform({a:0, b:1})
  if(u <= alpha)
    return sampler(n-1, func, prop, x_new, samples.concat(x_new))
  return sampler(n, func, prop, x_old, samples)
}
```

(b) [10P] Extend the program from task (a) such that you can visualize the histogram for $n = 20000$ samples, starting at $x_0 = 0.2$ with:

1. $f(x) = \frac{1}{2}e^{-\frac{(x-2)^2}{2}} + \frac{1}{2}e^{-\frac{(x+2)^2}{2}}$, with $g(x_{t+1}|x_t) \sim \mathcal{N}(x_t, 0.3)$.
2. $f(x) = \begin{cases} -0.1319x^4 + 1.132x^2 + 0.5, & x \in (-3,3) \\ 0, & \text{o.w.} \end{cases}$, with $g(x_{t+1}|x_t) \sim \mathcal{N}(x_t, 0.3)$

Solution:

```
var f1 = function(x){
  return 0.5 * Math.exp(-((x-2)*(x-2))/2) + 0.5 * Math.exp(-((x+2)*(x+2))/2)
}

var f2 = function(x){
  if(x < -3 || x > 3) return 0;
```

```
  return -0.1319 * Math.pow(x,4) + 1.132*x*x + 0.5
}

var prop = function(x){
  return gaussian({mu:x, sigma:0.3})
}

var n = 20000
var x_init = 0.2

viz.auto(sampler(n, f1, prop, x_init, []))
viz.auto(sampler(n, f2, prop, x_init, []))
```

(c) [5P] Describe what impact the parameter σ^2 of the proposal distribution has.

Solution: If σ^2 is too small: (high acceptance rate, low exploration) $\rightarrow$ slow convergence. If σ^2 is too large: (very low acceptance rate, high exploration) $\rightarrow$ slow convergence.

Exercise 2 (pGCL) **25P**

(a) [10P] Consider the following description of a stochastic process:

You live in the $\mathbb{Z}^2$ plane and are currently standing on position (x_0, y_0). You choose a direction (North, South, West, East) uniformly at random and take a step in this direction. You repeat this process until you have reached $(0,0)$.

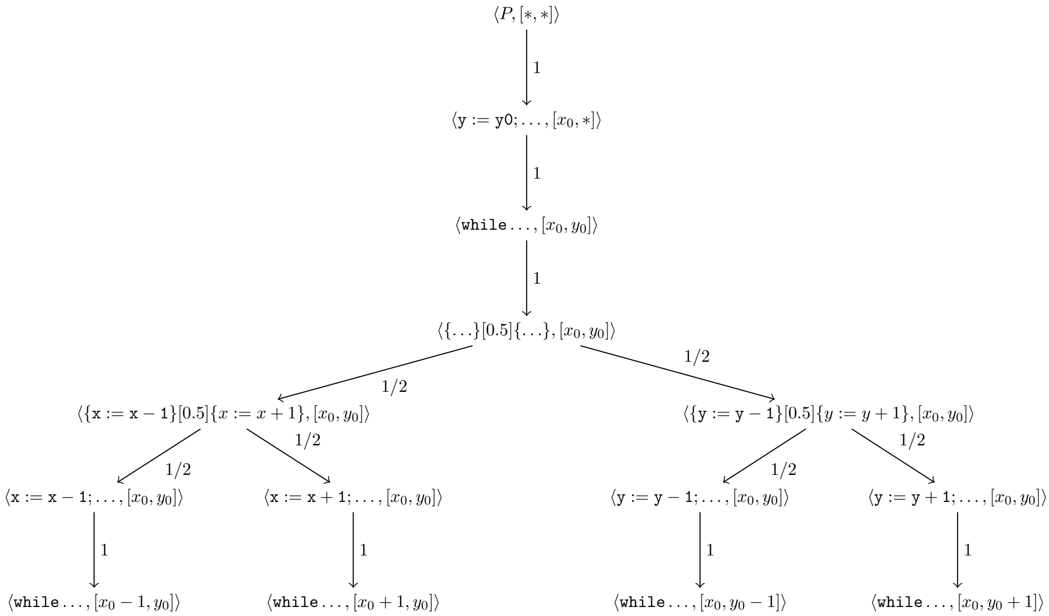
Design a pGCL program that exactly implements this behavior.

Solution:

```
x:=x0; y:=y0;
while(x != 0 || x != 0) {
  { {x:=x-1} [0.5] {x:=x+1} }
  [0.5]
  { {y:=y-1} [0.5] {y:=y+1} }
}
```

(b) [15P] Give the operational semantics of the pGCL program defined in (a) as a Markov chain for one loop iteration.

Solution: Notation: We write $[a, b]$ for the variable valuation $s(x) = a$, $s(x) = b$. If $x_0 \neq 0 \vee y_0 \neq 0$, then the initial fragment of the operational semantics Markov chain after one loop iteration looks like this:



If $x_0 = y_0 = 0$, then the loop body is never entered and the operational semantics is the following finite Markov chain:

$$\langle P, [*, *]\rangle \xrightarrow{1} \langle \texttt{y := y0}; \ldots, [x_0, *]\rangle \xrightarrow{1} \langle \texttt{while} \ldots, [x_0, y_0]\rangle \xrightarrow{1} \langle \downarrow, [0, 0]\rangle \xrightarrow{1} \langle \text{sink}\rangle \circlearrowleft 1$$

Exercise 3 (Markov chains with rewards) **25P**

(a) [15P] Consider a six-sided fair die. Compute the *expected* number of trials needed to see a six for the first time by modeling the problem with a Markov chain with rewards.

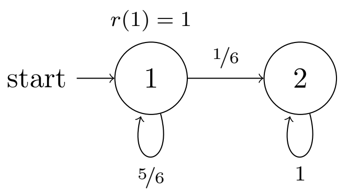
Hint: The following variant of the geometric series might be helpful: For $0 < q < 1$,

$$\sum_{k \geq 0} k q^k = \frac{q}{(1-q)^2} \; .$$

(b) [10P] Modify your reward Markov chain from part (a) such that it models the *conditional* expected number of trials needed to see a six for the first time *given that the first trial was no six*. Indicate the set F of "forbidden" states. You do not have to compute the conditional expected reward.

Solution:

(a) We use the following Markov chain with rewards and goal set $G = \{2\}$:



The idea is to count every trial (= visit to state 1) by assigning it a reward $r(1) = 1$. The expected number of trials is then given as $ER(1, \Diamond G)$. In order to compute this quantity, we consider all finite paths from 1 to $G = \{2\}$ in our reward Markov chain:

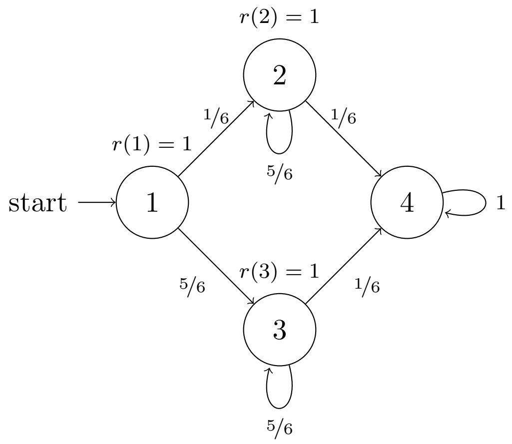
1 2
1 1 2
1 1 1 2
...

So all these paths are of the form $1^k\ 2$, where 1^k denotes the string $\underbrace{1\ldots1}_{k \text{ times}}$ and $k \geq 1$. Further, the reward of these paths is $r_G(1^k\ 2) = k$ and they occur with probability $Pr(1^k\ 2) = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6}$. Thus,

$$
\begin{aligned}
ER(1, \Diamond G) &= \sum_{k \geq 1} k \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \\
&= \frac{1}{6}\left(\sum_{k \geq 0} (k+1) \left(\frac{5}{6}\right)^{k}\right) \quad \text{(index shift)} \\
&= \frac{1}{6}\left(\sum_{k \geq 0} k \left(\frac{5}{6}\right)^{k} + \sum_{k \geq 0} \left(\frac{5}{6}\right)^{k}\right) \\
&= \frac{1}{6}\left(\frac{5/6}{(1-5/6)^2} + \frac{1}{1-5/6}\right) \quad \text{(formula from hint + standard geom. series)} \\
&= \frac{1}{6}\left(\frac{5/6}{1/6 \cdot 1/6} + 6\right) \\
&= 5 + 1 = 6\ .
\end{aligned}
$$

Remark: It is also possible (and easier) to compute the expected reward by solving a system of linear equations (see lecture).

(b) The following Markov chain "remembers" if the first trial was a six (state 2) or not (state 3):



To express that the first trial was no six, we let the forbidden states be $F = \{2\}$. The desired conditional expected reward is then $ER(1, \Diamond G \mid \neg\Diamond F)$, where $G = \{4\}$.

Exercise 4 (Complete lattices) $\boxed{\textbf{25P}}$

Let $(D, \sqsubseteq)$ be a complete lattice. Further, let $D \to D$ be the set of all monotone functions between elements of D. We lift the order $\sqsubseteq$ to the domain $D \to D$ by pointwise application, i.e., we define $(D \to D, \preceq)$ where $\preceq$ is given by:

$$f \preceq g \qquad \Longleftrightarrow \qquad \forall d \in D. \quad f(d) \sqsubseteq g(d).$$

(a) [5P] Show that $(D \to D, \preceq)$ is a partial order.

(b) [5P] Show that $(D \to D, \preceq)$ is a complete lattice.

(c) [15P] Now assume that D satisfies the *ascending chain condition*(ACC, for short), i.e., for all chains $S \subseteq D$, there exists a positive integer n such that $s_n = s_{n+1} = s_{n+2} = \dots$, or in other words *no infinite strictly ascending sequence of elements in D exists.*

Prove or disprove that fixed points of chains are continuous, i.e.,

$$\text{lfp}\left(\bigsqcup \mathcal{F}\right) \quad = \quad \bigsqcup \{\text{lfp}(f) \mid f \in \mathcal{F}\}$$

holds for all chains $\mathcal{F} \subseteq (D \to D)$.

Hint: If $(D, \sqsubseteq)$ is a complete lattice satisfying the ACC, $S \subseteq D$ a chain and $\Phi : D \to D$ is a monotone function, then $\Phi\left(\bigsqcup S\right) = \bigsqcup \Phi[S]$.

Solution:

(a) In the following let $f, g, h \in (D \to D)$.

- Reflexivity: By definition $f \preceq f$ holds if and only if for all $d \in D$, $f(d) \sqsubseteq f(d)$. This is trivial due to reflexivity of $(D, \sqsubseteq)$.

- Antisymmetry:

$$\begin{aligned}
& f \preceq g \ \wedge \ g \preceq f \\
\iff & \forall d \in D\colon \ f(d) \sqsubseteq g(d) \ \wedge \ g(d) \sqsubseteq f(d) \\
\iff & \forall d \in D\colon \ f(d) = g(d) \qquad\qquad \text{(because } \sqsubseteq \text{ is antisymmetric)} \\
\iff & f = d \ .
\end{aligned}$$

- Transitivity: Assume $f \preceq g \preceq h$. Then, for all $d \in D$, $f(d) \sqsubseteq g(d) \sqsubseteq h(d)$. Since $(D, \sqsubseteq)$ is transitive, we know for all $d \in D$ that $f(d) \sqsubseteq h(d)$. Hence, $f \preceq h$.

(b) Let $F \subseteq (D \to D)$ be an arbitrary subset of $(D \to D)$. We define $F_d = \{f(d) \mid f \in F\}$. Since $F_d \subseteq D$ and $(D, \sqsubseteq)$ is a complete lattice, $\bigsqcup F_d \in D$ exists. By definition, $f(d) \sqsubseteq \bigsqcup F_d$ holds for all $d \in D$ and $f \in F$. Hence, the function $g\colon D \to D,\ d \mapsto \bigsqcup F_d$ is the least upper bound of F and the greatest lower bound is obtained in a similar manner.

(c) Let $\mathcal{F} \subseteq (D \to D)$ be a chain. Also let

$$\begin{aligned}
h &= \bigsqcup \mathcal{F} \in (D \to D), \\
\tilde{f} &= \bigsqcup \{\mathrm{lfp}(f) \mid f \in \mathcal{F}\} \in D,
\end{aligned}$$

Case "$\sqsupseteq$": For every $f \in \mathcal{F}$, we have $f \preceq h$ by construction. Then (by a straightforward induction on n), $f^n \preceq h^n$ for each $n \geq 0$. Moreover, for each $f \in \mathcal{F}$, we have

$$\mathrm{lfp}(f) \ = \ \bigsqcup \left\{ f^n \left(\bigsqcup \emptyset \right) \mid n \geq 0 \right\} \ \sqsubseteq \ \bigsqcup \left\{ h^n \left(\bigsqcup \emptyset \right) \mid n \geq 0 \right\} \ = \ \mathrm{lfp}(h) \ ,$$

where $\emptyset$ denotes the empty chain in $(D \to D)$. Thus, $\tilde{f} \sqsubseteq \mathrm{lfp}(h)$.

Case "$\sqsubseteq$": We show that $h(\tilde{f}) \sqsubseteq \tilde{f}$. By Park's Lemma, this implies that $\mathrm{lfp}(h) \sqsubseteq \tilde{f}$.

$$\begin{aligned}
h(\tilde{f}) \ &= \ \bigsqcup \left\{ f(\tilde{f}) \mid f \in \mathcal{F} \right\} && \text{(Def. of } h\text{)} \\
&= \ \bigsqcup \left\{ f \left(\bigsqcup \{\mathrm{lfp}(g) \mid g \in \mathcal{F}\} \right) \mid f \in \mathcal{F} \right\} && \text{(Def. of } \tilde{f}\text{)} \\
&= \ \bigsqcup \left\{ \bigsqcup \{ f(\mathrm{lfp}(g)) \mid g \in \mathcal{F} \} \mid f \in \mathcal{F} \right\} && \text{(Using the Hint)} \\
&\sqsubseteq \ \bigsqcup \left\{ \bigsqcup \{\mathrm{lfp}(g) \mid g \in \mathcal{F}\} \mid f \in \mathcal{F} \right\} && (*) \\
&= \ \bigsqcup \left\{ \mathrm{lfp}(g) \mid g \in \mathcal{F} \right\} \ = \ \tilde{f} \ .
\end{aligned}$$

$(*)$ Since $\mathcal{F}$ is a chain we know that $f \sqsubseteq g$ or $g \sqsubseteq f$, for $f, g \in \mathcal{F}$.

- If $f \sqsubseteq g$ then $f(\mathrm{lfp}(g)) \ \sqsubseteq \ g(\mathrm{lfp}(g)) = \mathrm{lfp}(g)$.
- If $g \sqsubseteq f$ then
  $f(\mathrm{lfp}(g)) \ \sqsubseteq \ f\left(\bigsqcup \{g^n(\bigsqcup \emptyset) \mid n \geq 0\}\right) \ \sqsubseteq \ f\left(\bigsqcup \{f^n(\bigsqcup \emptyset) \mid n \geq 0\}\right) \ \sqsubseteq \ \mathrm{lfp}(f)$.

Thus, for each $f, g \in \mathcal{F}$ there exists an $h \in \{f, g\}$ such that $f(\mathrm{lfp}(g)) \ \sqsubseteq \ \mathrm{lfp}(h)$. Then every upper bound of $\{\mathrm{lfp}(h) \mid h \in \mathcal{F}\}$ is also an upper bound of $\{f(\mathrm{lfp}(g)) \mid g \in \mathcal{F}\}$ regardless of the choice of $f \in \mathcal{F}$.