Probabilistic Programming Homework 2

Submit your solutions to the TA in the homework submission box in the third floor of the E3-1 building by 2:00pm on 22 November 2017 (Wednesday). If you type up your solutions, you can email them to him (bskim90@kaist.ac.kr).

Question 1

This question is concerned with operational semantics for probabilistic programs. Throughout the question, we will use the definitions of value, evaluation context and the $\rightsquigarrow$ relation for the likelihood weighted importance sampler in the note "Implementing Inference Algorithm for Probabilistic Programs" in the course web page.

(a) In the lecture, we introduced `let` as macro that gets expanded as follows:

$$(\texttt{let}\ [x\ e]\ e') \equiv ((\texttt{fn}\ [x]\ e')\ e).$$

Suppose that we do not want to treat `let` as macro, but we want to regard it as one of the core constructs in the language. Extend the definitions of evaluation context C and the $\rightsquigarrow$ relation such that `let` expressions are evaluated directly by the $\rightsquigarrow$ relation, instead of first being macro-expanded and then being evaluated. Hint: You need to add one new case to the current definition of evaluation context C and one new rule for $\rightsquigarrow$.

(b) Consider the following program:

$$\begin{array}{l}(\texttt{let}\ [x\ (\texttt{sample}\ (\texttt{uniform-continuous}\ 0.1\ 0.9))] \\ \quad (\texttt{let}\ [y\ (\texttt{observe}\ (\texttt{flip}\ x)\ \texttt{true})] \\ \quad\quad (\texttt{if}\ (>\ x\ 0.5)\ 0\ 1)))\end{array}$$

Let us use e to denote this program. There are many ways of evaluating $(e, 1)$ to a value-weight pair using the $\rightsquigarrow$ relation extended in Part (a). For instance, the following repeated application of the the $\rightsquigarrow$ relation leads to the pair of value 1 and weight 0.2:

$$\begin{array}{l}\left(\begin{array}{l}(\texttt{let}\ [x\ (\texttt{sample}\ (\texttt{uniform-continuous}\ 0.1\ 0.9))] \\ \quad (\texttt{let}\ [y\ (\texttt{observe}\ (\texttt{flip}\ x)\ \texttt{true})] \\ \quad\quad (\texttt{if}\ (>\ x\ 0.5)\ 0\ 1)))\end{array}\ ,\ 1\right) \\ \rightsquigarrow \left(\begin{array}{l}(\texttt{let}\ [x\ 0.2] \\ \quad (\texttt{let}\ [y\ (\texttt{observe}\ (\texttt{flip}\ x)\ \texttt{true})]\ ,\ 1 \\ \quad\quad (\texttt{if}\ (>\ x\ 0.5)\ 0\ 1)))\end{array}\right) \\ \rightsquigarrow \left(\begin{array}{l}(\texttt{let}\ [y\ (\texttt{observe}\ (\texttt{flip}\ 0.2)\ \texttt{true})] \\ \quad (\texttt{if}\ (>\ 0.2\ 0.5)\ 0\ 1))\end{array}\ ,\ 1\right) \rightsquigarrow \left(\begin{array}{l}(\texttt{let}\ [y\ \texttt{true}] \\ \quad (\texttt{if}\ (>\ 0.2\ 0.5)\ 0\ 1))\end{array}\ ,\ 0.2\right) \\ \rightsquigarrow ((\texttt{if}\ (>\ 0.2\ 0.5)\ 0\ 1),\ 0.2) \rightsquigarrow ((\texttt{if}\ \texttt{false}\ 0\ 1),\ 0.2) \rightsquigarrow (1,\ 0.2)\end{array}$$

(Here I used an unexplained rule of $\rightsquigarrow$ for `let`) Using your rule for `let`, give two other possible execution sequences of the program e. Make it sure that the final values of your execution sequences are 0 and 1, respectively.

Question 2

The lightweight Metropolis-Hastings algorithm uses the following acceptance ratio:

$$\alpha((v_{n-1}, w_{n-1}, S_{n-1}), (v', w', S')) = \min\left(1, \frac{w' \cdot |S_{n-1}|}{w_{n-1} \cdot |S'|}\right).$$

Here $(v_{n-1}, w_{n-1}, S_{n-1})$ is the current state consisting of value v_{n-1}, weight w_{n-1}, and sequence of random choices S_{n-1}, and (v', w', S') is a newly proposed state. With the probability $\alpha((v_{n-1}, w_{n-1}, S_{n-1}), (v', w', S'))$, we accept this new state and set the next state (v_n, w_n, S_n) to be (v', w', S'). For detail, look at the note "Implementing Inference Algorithm for Probabilistic Programs" in the course web page.

Suppose that instead of the correct acceptance ratio above, we used the following variant:

$$\alpha'((v_{n-1}, w_{n-1}, S_{n-1}), (v', w', S')) = \min\left(1, \frac{w'}{w_{n-1}}\right).$$

Explain why this variant computes a wrong estimate for the following program:

```
(if (sample (flip 0.5))
  (or true (sample (flip 0.5)))
  false)
```

This (slightly silly) program does not contain any observe. Thus, its prior and posterior distributions coincide. In fact, the posterior distribution on boolean values assigns 0.5 to `true` and 0.5 to `false`. Hint: One way to answer this question is to focus on what the loop body of the variant of the lightweight MH algorithm does on v_n and to show that the target posterior distribution is not an invariant distribution (or stationary distribution) of the loop body.

Question 3

In the lectures, we learnt about the denotational semantics of a simple first-order programming language with discrete random choices. One big benefit of having such a semantics is that we can formally prove the equalities of programs; such equalities can form a theoretical basis for compiler optimisation. Proving the equation:

$$[\![\Gamma \vdash (\texttt{if false } e_1\ e_2) : t]\!] = [\![\Gamma \vdash e_2 : t]\!]$$

The proof is very similar to what was written in the lecture note. An optional question is to prove the following equation:

$$[\![\Gamma \vdash (\texttt{if } (\texttt{sample } (\texttt{flip } 0.7))\ \texttt{false true}) : \texttt{bool}]\!] = [\![\Gamma \vdash (\texttt{sample } (\texttt{flip } 0.3)) : \texttt{bool}]\!]$$

The proof of this equation is also similar to what you find in the lecture note. Checking how you need to modify the proof in the lecture will help you to understand what goes on there. Of course, proving this second equation is optional, and you don't have to do it.

Model answers

Answer to question 1

(a) One principle that can be used to solve this problem is to remember that ($\texttt{let}\ [x\ e]\ e'$) means (($\texttt{fn}\ [x]\ e'$) e), and to convert an appropriate evaluation context and an appropriate evaluation rule for (($\texttt{fn}\ [x]\ e'$) e) to those for `let`. We formed this model answer by following this principle.

Here is the new definition of evaluation context:

$$\begin{aligned} C ::= {} & [-] \\ \mid {} & (v_0\ v_1\ \ldots\ v_{m-1}\ C\ e_{m+1}\ \ldots\ e_{m+n}) \\ \mid {} & (\texttt{if}\ C\ e_1\ e_2) \\ \mid {} & (\texttt{let}\ [x\ C]\ e) \end{aligned}$$

We include ($\texttt{let}\ [x\ C]\ e$) because (($\texttt{fn}\ [x]\ e$) C) is an evaluation context.

We add the following rule in our definition of $\rightsquigarrow$:

$$\frac{}{C[(\texttt{let}\ [x\ v]\ e)],\ w \rightsquigarrow C[e[x := v]],\ w}$$

Any two legal execution sequences are acceptable as long as they lead to the final values 0 and 1.

Answer to question 2

Since the program does not contain any observe expressions, the weight obtained from its execution (by the $\rightsquigarrow$ relation) is always 1. This implies that

$$\alpha'((v_{n-1}, w_{n-1}, S_{n-1}), (v', w', S')) = 1$$

and that a proposed state always gets accepted.

Suppose that the current state $(v_{n-1}, w_{n-1}, S_{n-1})$ came from the true branch. Then, v_{n-1} is `true` and the first element of S_{n-1} is `true`. Now let's consider how the proposal behaves in this case. The proposal picks the first random choice in S_{n-1} with $1/2$ probability and the second with the same $1/2$ probability. Only when it picks the first random choice and changes it to `false`, the proposal goes to the false branch and proposes ($\texttt{false}, 1, [\texttt{false}]$). Otherwise, it goes to the true branch again, and proposes ($\texttt{true}, 1, [\texttt{true}, \texttt{true}]$) or ($\texttt{true}, 1, [\texttt{true}, \texttt{true}]$). Thus, the probabilities of all these possible outcomes are:

$$\begin{aligned} (\texttt{false}, 1, [\texttt{false}]) &: 1/2 \times 1/2 = 1/4 \\ (\texttt{true}, 1, [\texttt{true}, \texttt{true}]) &: 1/2 \times 1/2 \times 1/2 + 1/2 \times 1/2 = 1/8 + 1/4 = 3/8 \\ (\texttt{true}, 1, [\texttt{true}, \texttt{false}]) &: 1/2 \times 1/2 \times 1/2 + 1/2 \times 1/2 = 1/8 + 1/4 = 3/8 \end{aligned}$$

This means that $v_n = \texttt{false}$ with probability $1/4$ and $v_n = \texttt{true}$ with probability $3/4$:

$$k(v_n = \texttt{false} \mid v_{n-1} = \texttt{true}) = 1/4, \qquad k(v_n = \texttt{true} \mid v_{n-1} = \texttt{true}) = 3/4.$$

Suppose that the current state $(v_{n-1}, w_{n-1}, S_{n-1})$ came from the false branch. Then, v_{n-1} is `false` and the first element of S_{n-1} is `false`. By similar calculation as before, we can show that the proposal distribution proposes the following outcomes with specified probabilities:

$$\begin{aligned}(\texttt{false}, 1, [\texttt{false}]) &: 1/2\\ (\texttt{true}, 1, [\texttt{true}, \texttt{true}]) &: 1/2 \times 1/2 = 1/4\\ (\texttt{true}, 1, [\texttt{true}, \texttt{false}]) &: 1/2 \times 1/2 = 1/4\end{aligned}$$

This means that $v_n = \texttt{true}$ with probability $1/2$ and $v_n = \texttt{true}$ with probability $1/2$:

$$k(v_n = \texttt{false} \mid v_{n-1} = \texttt{false}) = 1/2, \qquad k(v_n = \texttt{true} \mid v_{n-1} = \texttt{false}) = 1/2.$$

Let p be our target distribution. That is,

$$p(\texttt{true}) = p(\texttt{false}) = 1/2.$$

Using what we have calculated so far, we can show that the target distribution p is not an invariant distribution of k.

$$\begin{aligned}&\sum_v k(v_n = \texttt{false} \mid v_{n-1} = v) \cdot p(v)\\ &= k(v_n = \texttt{false} \mid v_{n-1} = \texttt{true}) \cdot p(\texttt{true}) + k(v_n = \texttt{false} \mid v_{n-1} = \texttt{false}) \cdot p(\texttt{false})\\ &= 1/4 \cdot 1/2 + 1/2 \cdot 1/2\\ &= 3/8\\ &\neq 1/2 = p(\texttt{false}).\end{aligned}$$

Answer to question 3

The first equation is:

$$[\![\Gamma \vdash (\texttt{if false } e_1\ e_2) : t]\!] = [\![\Gamma \vdash e_2 : t]\!]. \tag{1}$$

To show this equality, pick an environment $\eta \in [\![\Gamma]\!]$. We will have to show that

$$[\![\Gamma \vdash (\texttt{if false } e_1\ e_2) : t]\!]\eta = [\![\Gamma \vdash e_2 : t]\!]\eta. \tag{2}$$

To do so, we will use the following important property between the δ function and the bind operator:

$$\delta_a \text{ bind } f = f(a). \tag{3}$$

This can be easily proved by using the definitions of δ_a and bind. Using this property, we prove the equation (2):

$$\begin{aligned}&[\![\Gamma \vdash (\texttt{if false } e_1\ e_2) : t]\!]\eta\\ &\quad= ([\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta) \text{ bind } ([b] \cdot [\![\Gamma \vdash e_1 : t]\!]\eta + [\neg b] \cdot [\![\Gamma \vdash e_2 : t]\!]\eta)\\ &\quad= \delta_{\text{ff}} \text{ bind } (\lambda b.\, [b] \cdot [\![\Gamma \vdash e_1 : t]\!]\eta + [\neg b] \cdot [\![\Gamma \vdash e_2 : t]\!]\eta)\\ &\quad= [\text{ff}] \cdot [\![\Gamma \vdash e_1 : t]\!]\eta + [\neg \text{tt}] \cdot [\![\Gamma \vdash e_2 : t]\!]\eta \qquad \text{by (3)}\\ &\quad= 0 \cdot [\![\Gamma \vdash e_1 : t]\!]\eta + 1 \cdot [\![\Gamma \vdash e_2 : t]\!]\eta\\ &\quad= [\![\Gamma \vdash e_2 : t]\!]\eta.\end{aligned}$$

The second equation is:

$$\begin{aligned}&[\![\Gamma \vdash (\texttt{if}\ (\texttt{sample}\ (\texttt{flip}\ 0.7))\ \texttt{false}\ \texttt{true}) : \texttt{bool}]\!] \\ &\quad = [\![\Gamma \vdash (\texttt{sample}\ (\texttt{flip}\ 0.3)) : \texttt{bool}]\!]\end{aligned} \tag{4}$$

We first calculate the semantics of

$$\Gamma \vdash (\texttt{flip}\ 0.7) : \texttt{dist[bool]} \quad \text{and} \quad \Gamma \vdash (\texttt{sample}\ (\texttt{flip}\ 0.7)) : \texttt{bool}.$$

Here are the outcomes of these calculations:

$$\begin{aligned}&[\![\Gamma \vdash (\texttt{flip}\ 0.7) : \texttt{dist[bool]}]\!]\eta \\ &\quad = ([\![\Gamma \vdash 0.7 : \texttt{rational}]\!]\eta)\ \text{bind}\ (\lambda r.\, \text{mean}(\texttt{flip})(r)) \\ &\quad = ([\![\Gamma \vdash 0.7 : \texttt{rational}]\!]\eta)\ \text{bind}\ \left(\lambda r.\, \delta_{[\text{tt}\mapsto r;\text{ff}\mapsto 1-r]}\right) \\ &\quad = \delta_{0.7}\ \text{bind}\ \left(\lambda r.\, \delta_{[\text{tt}\mapsto r;\text{ff}\mapsto 1-r]}\right) \\ &\quad = \delta_{[\text{tt}\mapsto 0.7;\text{ff}\mapsto 0.3]} \qquad \text{by (3).}\end{aligned}$$

$$\begin{aligned}&[\![\Gamma \vdash (\texttt{sample}\ (\texttt{flip}\ 0.7)) : \texttt{bool}]\!]\eta \\ &= ([\![\Gamma \vdash (\texttt{flip}\ 0.7) : \texttt{dist[bool]}]\!]\eta)\ \text{bind}\ (\lambda d.\, \text{mean}(\texttt{sample})(d)) \\ &= \delta_{[\text{tt}\mapsto 0.7;\text{ff}\mapsto 0.3]}\ \text{bind}\ (\lambda d.\, \text{mean}(\texttt{sample})(d)) \qquad \text{by the first cal.} \\ &= \delta_{[\text{tt}\mapsto 0.7;\text{ff}\mapsto 0.3]}\ \text{bind}\ (\lambda d.\, d) \\ &= [\text{tt} \mapsto 0.7; \text{ff} \mapsto 0.3] \qquad \text{by (3).}\end{aligned}$$

By similar calculations, we can also show:

$$[\![\Gamma \vdash (\texttt{flip}\ 0.3) : \texttt{dist[bool]}]\!]\eta = [\text{tt} \mapsto 0.3; \text{ff} \mapsto 0.7]$$

Using what we have so far, we prove our target equation (4): for all $v \in [\![\texttt{bool}]\!]$,

$$\begin{aligned}&[\![\Gamma \vdash (\texttt{if}\ (\texttt{sample}\ (\texttt{flip}\ 0.7))\ \texttt{false}\ \texttt{true}) : \texttt{bool}]\!]\eta(v) \\ &= \left(\begin{array}{l}([\![\Gamma \vdash (\texttt{sample}\ (\texttt{flip}\ 0.7)) : \texttt{bool}]\!]\eta)\ \text{bind}\ (\lambda b. \\ \quad [b] \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta + [\neg b] \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta)\end{array}\right)(v) \\ &= \left(\begin{array}{l}[\text{tt} \mapsto 0.7; \text{ff} \mapsto 0.3]\ \text{bind}\ (\lambda b. \\ \quad [b] \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta + [\neg b] \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta)\end{array}\right)(v) \\ &= \left(\begin{array}{l}0.7 \cdot ([\text{tt}] \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta(v) + [\neg \text{tt}] \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta(v)) \\ \quad + 0.3 \cdot ([\text{ff}] \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta(v) + [\neg \text{ff}] \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta(v))\end{array}\right) \\ &= \left(\begin{array}{l}0.7 \cdot (1 \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta(v) + 0 \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta(v)) \\ \quad + 0.3 \cdot (0 \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta(v) + 1 \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta(v))\end{array}\right) \\ &= 0.7 \cdot [\![\Gamma \vdash \texttt{false} : \texttt{bool}]\!]\eta(v) + 0.3 \cdot [\![\Gamma \vdash \texttt{true} : \texttt{bool}]\!]\eta(v) \\ &= 0.7 \cdot \delta_{\text{ff}}(v) + 0.3 \cdot \delta_{\text{tt}}(v) \\ &= [\text{tt} \mapsto 0.3; \text{ff} \mapsto 0.7](v) = [\![\Gamma \vdash (\texttt{sample}\ (\texttt{flip}\ 0.3)) : \texttt{bool}]\!]\eta(v).\end{aligned}$$