COMPSCI 687 Pop Quiz 3 — Solutions

Grading instructions. Each question is either correct or incorrect; there is no partial credit. Answers must be perfect down to the symbol, except for the accepted variants listed below. For Questions 2, 6, and 7, both the input/output sets and the function values are required; quantifiers may be included or omitted. In definitions, either $=$ or $:=$ is accepted. Periods at the end of answers are optional.

1. π is called a:

   Policy (Capitalization is not considered in grading).

2. Give the definition of π. Your answer should be two lines. The first specifies the input and output sets (e.g., $f : \mathcal{X} \to \mathcal{Y}$) and the second should specify the function values (e.g., $f(x) = \Pr(X = x)$). You may omit the quantifiers (e.g., $\forall x \in \mathcal{X}$).

$$\pi : \mathcal{S} \times \mathcal{A} \to [0, 1],$$
$$\pi(s, a) = \Pr(A_t = a \mid S_t = s).$$

3. Give the definition of G:

$$G = \sum_{t=0}^{\infty} \gamma^t R_t.$$

4. Give the definition of G_t:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k}.$$

   Accepted variants: Any other summation index may replace k if used consistently and without conflicting with t. For example, $G_t = \sum_{i=0}^{\infty} \gamma^i R_{t+i}$ is accepted.

5. Give the definition of H_t:

$$H_t = (S_0, A_0, R_0, \ldots, S_t, A_t, R_t).$$

   Accepted variants: Students may show any number of terms before the ellipsis, provided the terms follow the order $S_0, A_0, R_0, S_1, A_1, R_1, S_2, A_2, R_2, \ldots$. They need not finish a state–action–reward triple before the ellipsis. For example, both of the following are accepted:

$$H_t = (S_0, \ldots, S_t, A_t, R_t),$$
$$H_t = (S_0, A_0, R_0, S_1, A_1, R_1, S_2, \ldots, S_t, A_t, R_t).$$

   The history must end with R_t.

COMPSCI 687 Pop Quiz 3 — Solutions (continued)

6. Give the definition of J. Your answer should be two lines. The first specifies the input and output sets (e.g., $f : \mathcal{X} \to \mathcal{Y}$) and the second should specify the function values (e.g., $f(x) = \Pr(X = x)$). You may omit the quantifiers (e.g., $\forall x \in \mathcal{X}$).

$$J : \Pi \to \mathbb{R},$$
$$J(\pi) = \mathbf{E}[G; \pi].$$

Accepted variant: The second line may instead expand G:

$$J(\pi) = \mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t; \pi\right].$$

Both forms must include $; \pi$ inside the expectation. Here and elsewhere, we will not attempt to read whether a $\mathbb{R}$ or $\mathbf{E}$ is in blackboard bold or bold—any variant is acceptable, e.g., $\mathbb{R}, \mathrm{R}, \mathbf{R}$ and $\mathbb{E}, \mathbf{E}, \mathrm{E}$.

7. Give the definition of v^π. Your answer should be two lines. The first specifies the input and output sets (e.g., $f : \mathcal{X} \to \mathcal{Y}$) and the second should specify the function values (e.g., $f(x) = \Pr(X = x)$). You may omit the quantifiers (e.g., $\forall x \in \mathcal{X}$).

Required first line:

$$v^\pi : \mathcal{S} \to \mathbb{R}.$$

Accepted second lines: Any one of the following four forms is sufficient:

$$v^\pi(s) = \mathbf{E}[G_t \mid S_t = s; \pi],$$
$$v^\pi(s) = \mathbf{E}\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s; \pi\right],$$
$$v^\pi(s) = \mathbf{E}[G \mid S_0 = s; \pi],$$
$$v^\pi(s) = \mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \,\middle|\, S_0 = s; \pi\right].$$

In the second form, any other summation index may replace k consistently, as in Question 4. All four forms must include $; \pi$ inside the expectation. The return and conditioning time must match: G_t goes with $S_t = s$, whereas G goes with $S_0 = s$.

8. For the 6-state chain MDP used in lecture to define state-value functions, what is $p(s_\infty, a, s_\infty)$ equal to?

Any answer clearly indicating one will be accepted. E.g., students may write just "1" or $p(s_\infty, a, s_\infty) = 1$.