Stochastic Optimisation - Problem Sheet 4

Dom Hutchinson

November 29, 2020

Answer 1) i)

$$
\begin{aligned}
& \mathbb{P}^\pi(X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) \\
= \; & \mathbb{P}(X_0 = s_0) \prod_{k=1}^{t} \frac{\mathbb{P}^\pi(X_{0:k+1} = s_{0:k+1}, Y_{0:k} = a_{0:k})}{\mathbb{P}^\pi(X_{0:k} = s_{0:k}, Y_{0:k} = a_{0:k})} \cdot \frac{\mathbb{P}^\pi(X_{0:k+1} = s_{0:k}, Y_{0:k} = a_{0:k})}{\mathbb{P}^\pi(X_{0:k} = s_{0:k}, Y_{0:k-1} = a_{0:k-1})} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \mathbb{P}^\pi(X_{k+1} = s_{k+1} | X_{0:k} = s_{0:k}, Y_{0:k} = a_{0:k}) \cdot \mathbb{P}^\pi(Y_k = a_k | X_{0:k} = s_{0:k}, Y_{0:k-1} = a_{0:k-1}) \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \mathbb{P}^\pi(X_{k+1} = s_{k+1} | X_k = s_k, Y_k = a_k) \cdot \mathbb{P}^\pi(Y = a_k | X_k = s_k, Y_{k-1} = a_{k-1}) \text{ by Markov Property} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} p_k(s_{k+1} | s_k, a_k) q_k(a_k | s_{0:k}, a_{0:k-1})
\end{aligned}
$$

Answer 1) ii)

$$
\begin{aligned}
& \mathbb{P}^\pi(X_{0:t} = s_{0:k}) \\
= \; & \mathbb{P}(X_0 = s_0) \prod_{k=1}^{t} \mathbb{P}^\pi(X_k = s_k | X_{0:k-1} = s_{0:k-1}) \text{ by Bayes Rule} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \mathbb{P}^\pi(X_k = s_k | X_{k-1} = s_{k-1}) \text{ by Markov Property} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \left(\sum_{a \in A(s_k)} \mathbb{P}^\pi(X_k = s_k, Y_{k-1} = a | X_{k-1} = s_{k-1}) \right) \text{ by Marginalisation} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \left(\sum_{a \in A(s_k)} \mathbb{P}^\pi(X_k = s_k | Y_{k-1} = a, X_{k-1} = s_{k-1}) \mathbb{P}(Y_{k-1} a | X_{k-1} = s_{k-1}) \right) \text{ by Bayes Rule} \\
= \; & p_{X_0}(s_0) \prod_{k=1}^{t} \left(\sum_{a \in A(s_k)} p_k(s_k | s_{k-1}, a) q_k(a | s_{k-1}) \right)
\end{aligned}
$$

Answer 1) iii)

$$
\begin{aligned}
& \mathbb{P}^\pi(Y_{0:t} = a_{0:t}) \\
= \; & \sum_{s_{0:t} \in S^{t+1}} \mathbb{P}^\pi(X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) \text{ by Marginalisation} \\
= \; & \sum_{s_{0:t} \in S^{t+1}} p_{X_0}(s_0) \prod_{k=1}^{t} p_k(s_{k+1} | s_k, a_k) q_k(a_k | s_{0:k}, a_{0:k-1}) \text{ by } \texttt{1) i)}
\end{aligned}
$$

Answer 2) a.

To show that $\{X_t\}_{t\in T}$ is a markov chain it is sufficient to show that

$$\mathbb{P}(X_{t+1} = s_{t+1}|X_{0:t} = s_{0:t}) = \mathbb{P}(X_{t+1} = s_{t+1}|X_t = s_t)$$

By the definition of conditional probabilities we have

$$\mathbb{P}(X_{t+1} = s_{t+1}|X_{0:t} = s_{0:t}) = \frac{\mathbb{P}(X_{0:t+1} = s_{0:t+1})}{\mathbb{P}(X_{0:t} = s_{0:t})} \quad [1]$$

By marginalising we can re-express the numerator as

$$\mathbb{P}(X_{0:t+1} = s_{0:t+1}) = \sum_{a_{0:t}\in A^{t+1}} \mathbb{P}(X_{0:t+1} = s_{0:t+1}, Y_{0:t} = a_{0:t}) \quad [2]$$

Further, we can expand each term of this summation

$$\begin{aligned}
& \mathbb{P}(X_{0:t+1} = s_{0:t+1}, Y_{0:t} = a_{0:t}) \\
= \; & \frac{\mathbb{P}(X_{0:t+1} = s_{0:t+1}, Y_{0:t} = a_{0:t})}{\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t})} \cdot \frac{\mathbb{P}(X_{0:t} = s_{0:t} = Y_{0:t} = a_{0:t})}{\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})} \cdot \mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}) \\
= \; & \mathbb{P}(X_{t+1} = s_{t+1}|X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t})\mathbb{P}(Y_t = a_t|X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}) \\
= \; & \mathbb{P}(X_{t+1} = s_{t+1}|X_t = s_t, Y_t = a_t)\mathbb{P}(Y_t = a_t|X_t = s_t, Y_{t-1} = a_{t-1})\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}) \\
= \; & p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})
\end{aligned}$$

where the penultimate step is due to the use of Markovian randomised decision rules. Substituting this expression back into [2] gives the following expression

$$\mathbb{P}(X_{0:t+1} = s_{0:t+1}) = \sum_{a_{0:t}\in A^{t+1}} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})$$

This expression can be restated as a recursive expression

$$\begin{aligned}
& \mathbb{P}(X_{0:t+1} = s_{0:t+1}) \\
= \; & \sum_{a_{0:t}\in A^{t+1}} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}) \\
= \; & \sum_{a_t\in A}\sum_{a_{0:t-1}\in A^t} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}) \\
= \; & \left(\sum_{a_t\in A} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\right)\left(\sum_{a_{0:t-1}\in A^t} \mathbb{P}(X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})\right) \\
= \; & \left(\sum_{a_t\in A} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)\right)\mathbb{P}(X_{0:t} = s_{0:t})
\end{aligned}$$

Substituting this expression into [1] we get

$$\begin{aligned}
\mathbb{P}(X_{t+1} = s_{t+1}|X_{0:t} = s_{0:t}) &= \frac{\mathbb{P}(X_{0:t+1} = s_{0:t+1})}{\mathbb{P}(X_{0:t} = s_{0:t})} \\
&= \frac{\sum_{a_t\in A} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)}{\mathbb{P}(X_{0:t} = s_{0:t})} \\
&= \sum_{a_t\in A} p_t(s_{t+1}|s_t, a_t)q_t(a_t|s_t)
\end{aligned}$$

Since the RHS of this final expression is independent of $s_{0:t-1}$, $\{X_t\}_{t\in T}$ is a Markov chain.

Answer 2) b.
From `2) a.` we deduced that the transition kernel of our process is

$$\mathbb{P}(X_{t+1} = s' | X_t = s) = \sum_{a \in A} p_t(s'|s, a) q_t(a|s)$$

Since the transition and decision probabilities are stationary we have that $p(\cdot|\cdot, \cdot) = p_t$ and $q(\cdot|\cdot) = q_t(\cdot|\cdot)$ for all

Answer 3)
For ease of notation let $p_{Y_t}(a) := \mathbb{P}(Y_t = a | X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1})$.
In order to show that when the agent uses the procedure defined in the question they do indeed implement the decision rule with decision function $q_t(a|s_{0:t}, a_{0:t-1})$, I shall show that

$$p_{Y_t}(a) = q_t(a|s_{0:t}, a_{0:t-1})$$

By the definition of Y_t, we have

$$p_{Y_t}(a) = \mathbb{P}\big(\Phi_t(U_t|s_{0:t}, a_{0:t-1}) = a | X_{0:t} = s_{0:t}, Y_{0:t-1} = a_{0:t-1}\big)$$

By the definition of $\Phi_t(\cdot)$, we have three separate cases

$$p_{Y_t}(a) = \begin{cases} \mathbb{P}(U_t \leq q_t(a(1)|s_{0:t}, a_{0:t-1})) & \text{if } a = a(1) \\ \mathbb{P}\left(U_t \in \left(\sum_{j=1}^{k-1} q_t(a(j)|s_{0:t}, a_{0:t-1}), \sum_{j=1}^{k} q_t(a(j)|s_{0:t}, a_{0:t-1})\right]\right) & \text{if } a = a(k) \text{ with } k \in [2, M-1] \\ \mathbb{P}\left(U_t > \sum_{j=1}^{M-1} q_t(a(j)|s_{0:t}, a_{0:t-1})\right) & \text{if } a = a(M) \end{cases}$$

Since U_t is a uniform distribution on $[0, 1]$ and $q_t(\cdot)$ is a probability distribution, we can restate these cases as

$$p_{Y_t}(a) = \begin{cases} q_t(a(1)|s_{0:t}, a_{0:t-1}) & \text{if } a = a(1) \\ \sum_{j=1}^{k} q_t(a(j)|s_{0:t}, a_{0:t-1}) - \sum_{j=1}^{k-1} q_t(a(j)|s_{0:t}, a_{0:t-1}) & \text{if } a = a(k) \text{ with } k \in [2, M-1] \\ 1 - \sum_{j=1}^{M-1} q_t(a(j)|s_{0:t}, a_{0:t-1})) & \text{if } a = a(M) \end{cases}$$

By the fact that $q_t(\cdot)$ is a probability distribution, and the definition of summation, the cases can be further simplified to

$$p_{Y_t}(a) = \begin{cases} q_t(a(1)|s_{0:t}, a_{0:t-1}) & \text{if } a = a(1) \\ q_t(a(k)|s_{0:t}, a_{0:t-1}) & \text{if } a = a(k) \text{ with } k \in [2, M-1] \\ q_t(a(M)|s_{0:t}, a_{0:t-1}) & \text{if } a = a(M) \end{cases}$$

This final set of cases are just a partition of the decision function, and thus can be simplified to

$$p_{Y_t}(a) = q_t(a|s_{0:t}, a_{0:t-1}) \ \forall \ a \in A$$

□