Application Notes - Stochastic Optimisation

Dom Hutchinson

November 24, 2020

Question 1)
Consider the following stochastic system. Let $T := \{0, \dots, N-1\}$ be a finite time-horizon, $X_t \in S$ be the system state at epoch $t \in T$, $Y_t \in A$ be the action taken at epoch $t \in T$, $A(s) \subseteq A$ be the admissible actions when in state $s \in S$. The stochastic system has the follow dynamics

$$\begin{array}{rcl} \Psi_t & : & S \times A \times B \to S \\ X_{t+1} & = & \Psi_t(X_t, Y_t, U_t) \\ \Phi_t & : & S \times C \to A \\ Y_{t+1} & = & \Phi_t(X_t, V_t) \\ R_t & : & S \times A \times D \to \mathbb{R} \\ & & \mathcal{R}_t(X_t, Y_t, W_t) \end{array}$$

where $U_t \sim \text{Uni}(B),\ U_t \sim \text{Uni}(C),\ W_t \sim \text{Uni}(D)$ for some discrete systems B, C, D. Assume that $X_0, \{U_t\}_{t\in T}, \{V_t\}_{t\in T}, \{W_t\}_{t\in T}$ are all mutually independent.

The objective of this task is to maximised the total expected reward from this system

$$\max \mathbb{E}\left[\sum_{t\in T} R_t(X_t, Y_t, W_t)\right]$$

Question 1) (a)
Show that the problem of maximising the expected total reward for this stochastic system is equivalent to the Markov Decision Problem.

Answer 1) (a)
This requires us to show two properties

i). This stochastic system exhibits Markovian Dynamics

$$\mathbb{P}(X_{t+1} = s_{t+1} | X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) = \mathbb{P}(X_{t+1} = s_{t+1} | X_t = s_t, Y_t = a_t)$$

ii). The expected total reward admits the following representation

$$\mathbb{E}\left[\sum_{t\in T} R_t(X_t, Y_t, W_t)\right] = \mathbb{E}\left[\sum_{t\in T} r_t(X_t, Y_t)\right]$$

At epoch $t = 1$ we have

$$\begin{array}{rcl} X_1 & = & \Psi_t(X_0, Y-0, U_0) \\ & = & \Psi_1(X_0, \Phi_0(X_0, V_0), U_0) \\ \implies X_1 & = & \tilde{\Psi}_1(X_0, U_0, V_0) \end{array}$$

for a new function $\tilde{\Psi}_1 : S \times B \times C \to S$. Also, at epoch $t = 1$ we have

$$\begin{array}{rcl} Y_1 &=& \Phi_1(X_1, V_1) \\ &=& \Phi_1(\tilde{\Psi}_1(X_0, U_0, V_0), V_1) \\ \implies Y_1 &=& \tilde{\Phi}_1(X_0, U_0, V_{0:1}) \end{array}$$

for a new function $\tilde{\Phi}_1 : S \times B \times C^2 \to A$. We can extend this to the general epoch t

$$\begin{array}{rcl} X_t &=& \tilde{\Psi}_t(X_0, U_{0:t-1}, V_{0:t-1}) \\ Y_t &=& \tilde{\Phi}_t(X_0, U_{0:t-1}, V_{0:t}) \end{array}$$

where our general mapping functions have signatures

$$\begin{array}{rcl} \tilde{\Psi}_t &:& S \times B^t \times C^t \to S \\ \tilde{\Phi}_t &:& S \times B^t \times C^{t+1} \to A \end{array}$$

As we are allowed to assume that $X_0, \{U_t\}_{t\in T}, \{V_t\}_{t\in T}, \{W_t\}_{t\in T}$ are all mutually independent. We have that U_t & $(X_{0:t}, Y_{0;t})$ are mutually independent and W_t & (X_t, Y_t) are mutually independent. [1]

Consider the transition probabilities

$$\begin{array}{rl} & \mathbb{P}(X_{t+1} = s_{t+1} | X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) \\ = & \mathbb{P}(\Psi_t(X_t, Y_t, U_t) = s_{t+1} | X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) \text{ by def. } X_{t+1} \\ = & \mathbb{P}(\Psi_t(s_t, a_t, U_t) = s_{t+1} | X_{0:t} = s_{0:t}, Y_{0:t} = a_{0:t}) \text{ by conditions} \\ = & \mathbb{P}(\Psi_t(s_t, a_t, U_t) = s_{t+1}) \text{ as } U_t \perp\!\!\!\perp (X_{0:t}, Y_{0:t}) \\ = & \mathbb{P}(X_{t+1} = s_{t+1} | X_t = s_t, Y_t = a_t) \end{array}$$

This shows that the stochastic system exhibits markovian dynamics.

Question 1) (b)
Identify the elements of the equivalent Markov Decision Problem.

Answer 1) (b)
This requires us to identify the following

i). Transition probabilities

$$p_t(s'|s, a) := \mathbb{P}(X_{t+1} = s | X_t = s, Y_t = a)$$

ii). Equivalent reward

$$r_t(s, a)$$

We derive the transition probabilities as follows

$$\begin{array}{rcl} p_t(s'|s, a) &:=& \mathbb{P}(X_{t+1} = s' | X_t = s, Y_t = a) \\ &=& \mathbb{P}(\Psi_t(X_t, Y_t, U_t) = s' | X_t = s, Y + t = a) \text{ by def. } X_{t+1} \\ &=& \mathbb{P}(\Psi_t(s, a, U_t) = s') \text{ by conditions} \\ &=& \mathbb{E}\left[\mathbb{1}\{\Psi_t(s, a, U_t) = s'\}\right] \\ &=& \sum_{u\in B} \mathbb{1}\{\Psi_t(s, a, u) = s'\} \cdot f_{U_t}(u) \end{array}$$

[1] Proof is long and given in slides

We have

$$\begin{array}{rcl}\mathbb{E}\left[\sum_{t\in T}R_t(X_t,Y_t,W_t)\right] & = & \sum_{t\in T}\mathbb{E}\left[R_t(X_t,Y_t,W_t)\right] \\ & = & \sum_{t\in T}\mathbb{E}\left[\mathbb{E}\left[R_t(X_t,Y_t,W_t)|X_t,Y_t\right]\right] \text{ by Tower Property}\end{array}$$

Define $r_t(s,a) := \mathbb{E}\left[R_t(X_t,Y_t,W_t)|X_t=s,Y_t=a\right]$. This gives us a representation for expected total reward

$$\mathbb{E}\left[\sum_{t\in T}R_t(X_t,Y_t,W_t)\right] = \mathbb{E}\left[\sum_{t\in T}r_t(X_t,Y_t)\right]$$

Since W_t & (X_t,Y_T) are mutually independent we can get a deterministic expression for $r_t(s,a)$

$$\begin{array}{rcl}r_t(s,a) & = & \mathbb{E}\left[R_t(X_t,Y_t,W_t)|X_t=s,Y_t=a\right] \\ & = & \mathbb{E}\left[R_t(s,a,W_t)\right] \text{ by conditions} \\ & = & \sum_{w\in D}R_t(S,a,w)f_{W_t}(w) \text{ by def. expectation}\end{array}$$