Class Tutorial 10

1. Stochastic Approximation ODE Approach

Consider an initially empty urn to which balls, either red or black, are added one at a time. Let y_n denote the *number* of red balls at time n, and $\theta_n \triangleq y_n / n$ the *fraction* of red balls at time n.

The probability of adding a red ball at time $n+1$ is a function of θ_n alone, and we denote it by $p(\theta_n)$. We are interested in the limit $\lim_{n\to\infty}\theta_n$. Let ξ_{n+1} denote the following random variable

$$\xi_{n+1} = \begin{cases} 1 \quad , & (n+1)\text{st ball is red} \\ 0 \quad , & (n+1)\text{st ball is black} \end{cases}$$

a. Write an update equation for θ_n in the form of a stochastic approximation $\theta_{n+1} = \theta_n + a_n\left[h(\theta_n) + \omega_n\right]$, where ω_n is a martingale difference noise.

b. Write down the corresponding ODE. What are the asymptotically stable equilibria?

c. Assume θ^* is a unique stable equilibrium of the ODE, and that $p(\theta)$ is Lipschitz. Show that $\lim_{n\to\infty}\theta_n = \theta^*$.

Solution

a. We have $y_{n+1} = y_n + \xi_{n+1}$, therefore

$$\begin{aligned} \theta_{n+1} &= y_{n+1}/(n+1) = \\ &= \frac{y_n + \xi_{n+1}}{n+1} = \frac{n y_n}{n(n+1)} + \frac{\xi_{n+1}}{n+1} = \\ &= \frac{n}{n+1}\theta_n + \frac{\xi_{n+1}}{n+1} = \theta_n + \frac{1}{n+1}(\xi_{n+1} - \theta_n) \end{aligned}$$

We now transform it to a stochastic approximation form:

$$\begin{aligned} \theta_{n+1} &= \theta_n + \frac{1}{n+1}\left(\xi_{n+1} + \mathrm{E}(\xi_{n+1} \mid \theta_n) - \mathrm{E}(\xi_{n+1} \mid \theta_n) - \theta_n\right) \\ &= \theta_n + \frac{1}{n+1}\left(p(\theta_n) - \theta_n + \xi_{n+1} - p(\theta_n)\right) \end{aligned}$$

Here, $\omega_n \equiv \xi_{n+1} - p(\theta_n)$, and obviously $\mathrm{E}(\omega_n \mid \theta_n) = 0$. Also, $\alpha_n = \dfrac{1}{n+1}$ and $h(\theta_n) = p(\theta_n) - \theta_n$.

b. $\dot{\theta}(t) = h(\theta(t)) = p(\theta(t)) - \theta(t)$. The equilibria satisfy $p(\theta^*) = \theta^*$. Stability criterion: by linearization around θ^*:

$$p'(\theta^*) - 1 < 0$$
$$p'(\theta^*) < 1$$

c. We will use the stochastic approximation convergence theorem (Theorem 1 in the lecture notes). We need to show that:

Assumption G1: $\sum_n a_n = \infty,\ \sum_n a_n^{\ 2} < \infty$, clearly holds for $\alpha_n = \dfrac{1}{n+1}$.

Assumption N1: ω_n is a martingale difference, and $\mathrm{E}\left[\|\omega_n\|^2 \middle| \mathcal{F}_n\right] \le A + B\|\theta_n\|^2$. Clearly holds since the noise is bounded.

The sequence θ_n is bounded w.p. 1. Here this holds by definition of θ_n.

2.Stochastic Approximation Contraction Mapping Approach

Consider the Q-learning algorithm where π is some ‘explorative’ policy.

a) Show that the updating equations of the Asynchronous Q-learning algorithm can be written as a sum of two terms, a $TQ_n - Q_n$ and a martingale noise.
b) Assume that the step-size requirements satisfy the assumption

$$\forall s, a\ (w.p\ 1)\ \sum \alpha_n(s,a) = \infty, \sum \alpha_n^2(s,a) < \infty,$$

Prove the convergence of Q-learning to the optimal Q-function.

Q-learning
Initialize: $Q_0(s,a)$ **For** k=0,1,... **do** Sample $a_n \sim \pi(s_n)$ Sample $s_{n+1} \sim P(\cdot \mid s_n, a_n), r(s_n, a_n, s_{n+1})$ $\delta_n = r(s_n, a_n, s_{n+1}) + \gamma \max_{a'} Q_n(s_{n+1}, a') - Q_n(s_n, a_n)$ $Q_{n+1}(s_n, a_n) = Q_n(s_n, a_n) + \alpha_n(s_n, a_n)\delta_n$

Solution

a) We consider updating the full vector of the Q-function. For all the s, a entries which are not the current state-action pair we set

$$\alpha_n(s_n, a_n) = 0.$$

Furthermore, we have that

$$\delta_n = (TQ_n)(s_n, a_n) - Q_n(s_n, a_n) + \omega_n$$

Where

$$(TQ_n)(s_n, a_n) = E_{s' \sim P(\cdot|s_n, a_n)}\left[r(s_n, a_n, s') + \gamma \max_{a'} Q_n(s', a')\right]$$
$$\omega_n = r(s_n, a_n, s_{n+1}) + \gamma \max_{a'} Q_n(s_{n+1}, a') - (TQ_n)(s_n, a_n)$$

It is easy to see that ω_n is a martingale noise by verifying that

$$E[\omega_n \mid \mathcal{F}_n] = 0$$

Where $\mathcal{F}_n$ is the entire history, meaning the value of Q_n ,the current state and action, s_n, a_n, and $\alpha_n(s_n, a_n)$.

b) We use Theorem 9.6 from the lectures to show the convergence. We start by observing that T is a contraction operator. Furthermore,

- The step-size requirements holds by assumption.
- As we saw, ω_n is a martingale noise. It is also satisfied that

$$E[\omega_n^2 \mid \mathcal{F}_n] \leq |Q|_\infty^2$$

Thus, using Theorem 9.6 we get that Q-learning converges w.p 1 to the optimal Q function.