1 Probabilistic or Stochastic DP-Finite Horizon continued

a) The next state is certain but the reward/cost obtained in the current state is stochastic,

$$f_t(i) = \min_x [Expected\ Cost\ c(i,x) + f_{t+1}(j)], \tag{1}$$

where

$$c(i,x) = \sum_j p(i,x,j)c(i,x,j) \tag{2}$$

b) The next state is uncertain and the reward/cost obtained in the current state is stochastic. In general

$$f_t(i) = \min_x [Expected\ Cost\ c(i,x) + \sum_j p(i,x,j)f_{t+1}(j)], \tag{3}$$

c) There are problems where $c(i,x)$ is fixed (not an expected value) and the next state is uncertain.

$$f_t(i) = \min_x [c(i,x) + \sum_j p(i,x,j)f_{t+1}(j)], \tag{4}$$

d) There are also problems where $c(i,x)$ does not exist. The next state is obviously uncertain.

$$f_t(i) = \min_x [\sum_j p(i,x,j)f_{t+1}(j)], \tag{5}$$

Stochastic DP (finite horizon) is also represented as decision trees.

1.1 Example 1

ABC has D dollars to allocate for drilling sites $1, 2, \cdots, T$. If $\$x$ is allocated to site t then the probability of finding oil is $P_t(x)$. The oil at site t is worth $\$r_t$. Formulate DP to maximize the expected value of all oil found at sites $1, 2, \cdots, T$. [1]

1.2 Example 2

A town must decide how many bass to catch and sell in a year. Each bass in year t is sold for $\$s_t$. The cost of catching x bass is $c_t(x, b)$ where b is the number of bass in the lake at the beginning of the year. Bass reproduce and the number of bass in the lake at the beginning of a year is d times more than the number left at the end of the previous year, where d is a random number that follows the probability distribution $P(D = d) = p(d)$. Formulate a DP recursion to maximize revenue over T years. [1]

1.3 Example 3

When Sally arrives at a bank there are 30 minutes of lunch time remaining. If she joins/enters the head of the line and enters service before 30 minutes then she earns a reward r. She does not enjoy waiting in line and incurs a cost of c for each minute she waits. During a minute when n people are waiting ahead of Sally, there is a probability $p(x|n)$ that x people will complete transactions. Formulate DP to maximize her expected net revenue (reward-waiting cost). She can decide not to join and leave with 0 cost. She can also decide to leave at a later time if she finds that its not worth waiting again by incurrinng 0 cost for that minute.

2 Markov Chanis

References

[1] W. L. Winston. *Introduction to Mathematical Programming, Vol 1.* Thompson, 2003.