Announcements

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

- In-Class Assignment from last class and Assignment 3 are is due now
- Assignment 4 on Canvas – due next Tuesday
- Homework assignment 2 is graded:
  - This time I graded question 2
  - Median was about 8/10
  - Most common causes of lost points: no model description, i.e., not definition of states, actions, rewards, optimality equations. Solution without using DP.

Last Time

- Policy evaluation
- Optimal solution to Markov Decision Processes

Today

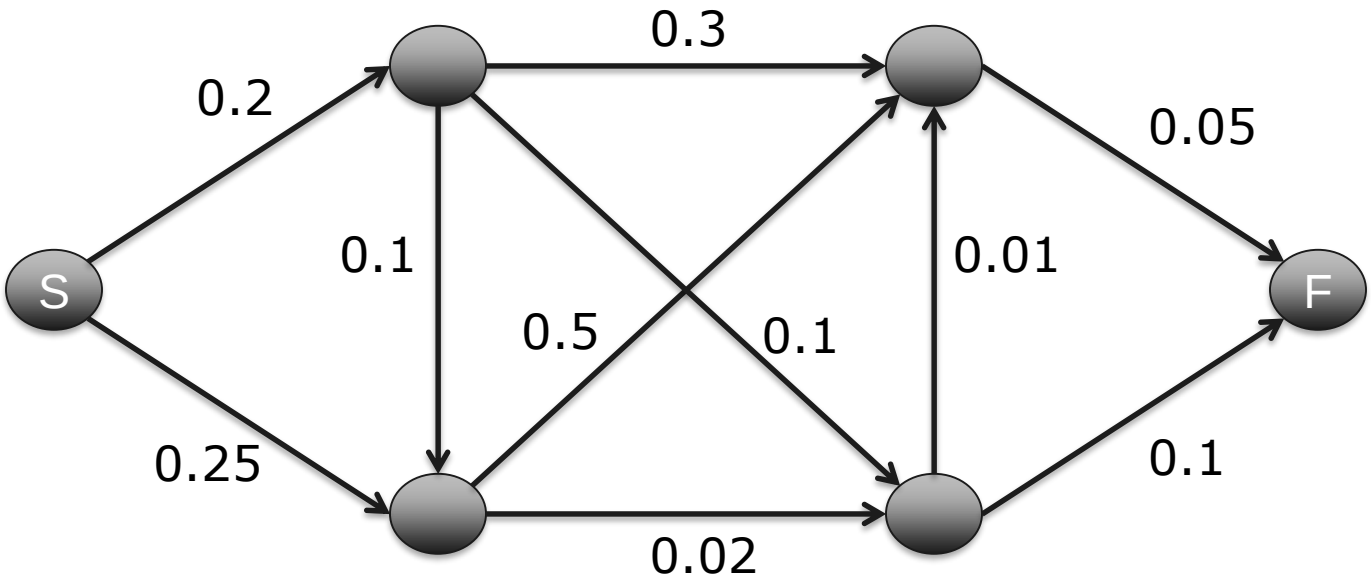
Examples:

- Zombies
- Inventory control

Zombie Avoidance

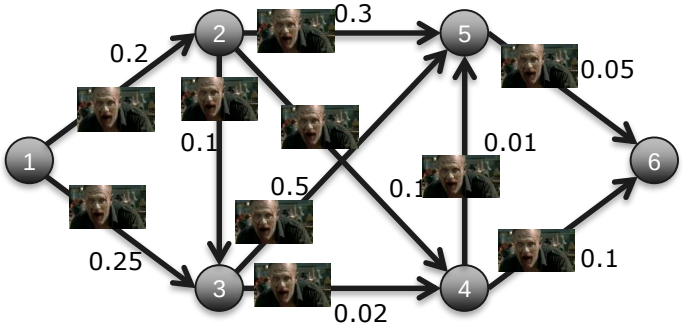
Formulate and solve a DP for the following problem:

- You must traverse the following network in which edge weights are the probability of encountering a zombie
- If you encounter a zombie along an edge then at the next vertex you randomly select an edge with equal probability; otherwise you select deterministically.
- Your goal is to minimize encounters



Zombie Avoidance

- What is the goal of solving the problem?
- What will the optimal policy look like?
- What is the state?
- What does the value function represent?



Optimality Equations
Define the value function as the probability of not encountering a zombie

Boundary Condition: $v_6 = 1$

Stage 5: $v_5 = 0.95v_6 = .95$

Stage 4: $v_4 = \max_{\{5,6\}}\{0.99v_5, 0.9v_6\} = .94$

Stage 3: $v_3 = \max_{\{4,5\}}\{0.98v_4, 0.5v_5\} = .92$

Stage 2: $v_2 = \max_{\{3,4,5\}}\{0.90v_3, 0.9v_4, 0.7v_5\} = .85$

Stage 1: $v_1 = \max_{\{2,3\}}\{0.8v_2, 0.75v_3\} = .69$

Optimal path is: 1→ 3→4→5→ 6, probability of no encounter on this path = 0.69

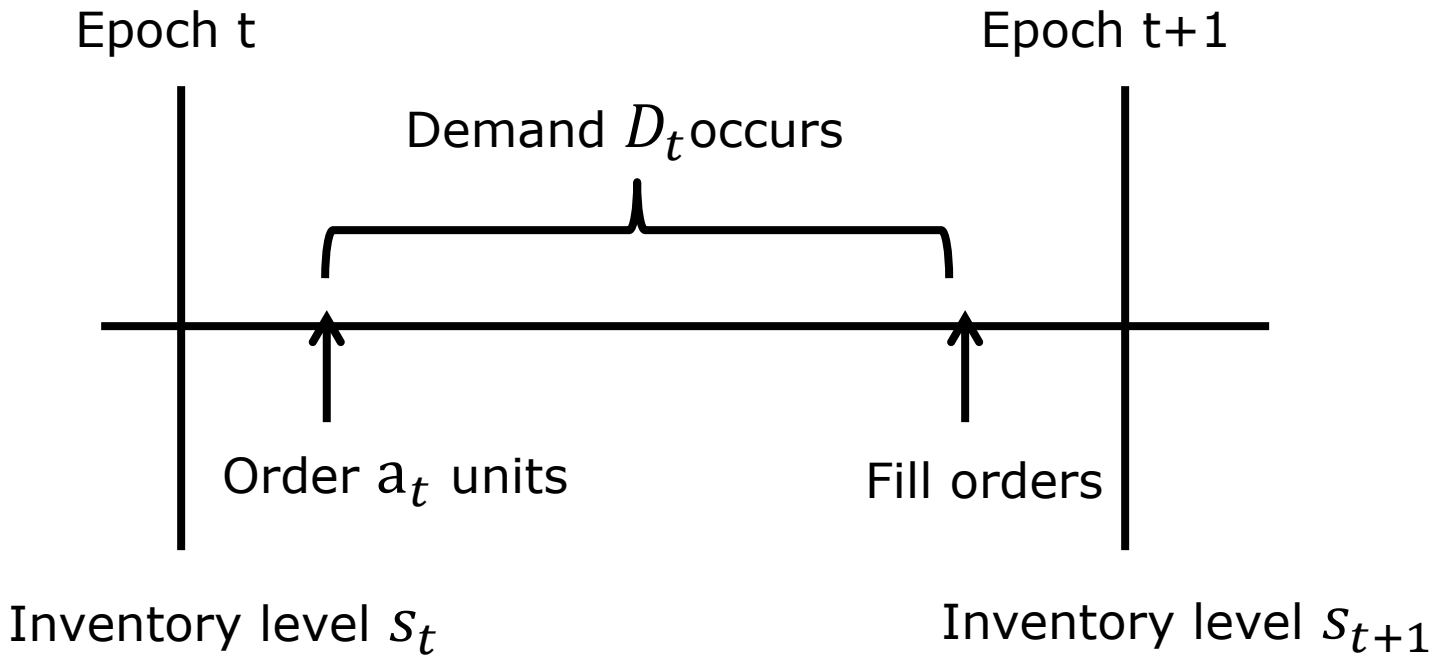
Inventory Control

The lotsizing problem can be extended to the case of uncertain demand

- A Markov chain is used to represent demand uncertainty
- Decision maker chooses how much to produce at each decision epoch
- If inventory exceeds demand an inventory cost is incurred

Inventory Control

The timing of events for the decision process:



Demand that is not filled is lost.

Inventory Control MDP

Decision epochs: $t = 1,2,\ldots.,T$.

Actions: order quantity: $\mathrm{a_t} \in \{0,1,2,\ldots,\mathrm{M} - \mathrm{s_t}\}$, where M is max available space for inventory

States: the state is the inventory level, defined by the following transfer equation:

$$\mathrm{s_{t+1}} = \max\{s_t + a_t - D_t, 0\} = (s_t + a_t - D_t)^+$$

States are uncertain because demand, D_t, is a random variable.

Transition Probabilities: $\mathrm{p_t}(j|s_t, a_t) = \begin{cases} 0, & j > s_t + a_t \\ p_{s_t+a_t-j}, & if\ s_t + a_t \geq j > 0 \\ q_{s_t+a_t}, & if\ \ j = 0 \end{cases}$

where $q_{s_t+a_t}$ is the probability demand exceeds supply.

Inventory Control MDP

Rewards are expected profit which includes expected revenue (p per unit) minus production cost and inventory cost.

Expected revenue: $E_{D_t}[p \times min(s_t + a_t, D_t)]$

Production cost (fixed + variable): $\mathrm{K} \times \mathrm{I}(\mathrm{a_t}) + \mathrm{c}(\mathrm{a_t})$, where $I(a_t) = \begin{cases} 0 \ if \ a_t = 0 \\ 1 \ if \ a_t > 0 \end{cases}$

Inventory cost: $\mathrm{h} \times (\mathrm{s_t} + \mathrm{a_t})$

Rewards:

$$\mathrm{r_t}(s_t, a_t) = E_{D_t}[p \times min(s_t + a_t, D_t)] - \mathrm{K} \times \mathrm{I}(\mathrm{a_t}) + \mathrm{c}(\mathrm{a_t}) - \mathrm{h} \times (\mathrm{s_t} + \mathrm{a_t}), \forall (s_t, a_t)$$

$$r_T(s_T) = R(s_T), \forall s_T$$

Example

Consider the following specific example of the inventory control MDP:

Assume $\mathrm{K} = 4,\ \mathrm{c}(\mathrm{a_t}) = 2a_t,\ R(s_T) = 0, \forall s_T,\ \ h(x) = x,\ M = 3,\ T = 3,$ $p = 8$, and demand distribution

$$p_d = \begin{cases} 0.25, & if\ d = 0 \\ 0.5, & if\ d = 1 \\ 0.25, & if\ \ d = 2 \end{cases}$$

What is the expected profit if zero stock is available at the first stage? What is the optimal ordering policy?

Solution

States: s_t= inventory level at start of epoch t

Actions: a_t = order quantity during epoch t

Rewards: $r_t(s_t, a_t)$ can be expressed as an $|S| \times |A|$ matrix as follows

$$\text{Reward Matrix} = \begin{bmatrix} 0 & -1 & -2 & -5 \\ 5 & 0 & -3 & X \\ 6 & -1 & X & X \\ 5 & X & X & X \end{bmatrix}$$

Transition Probability Matrix (example for $a_t = 0$):

$$P_t(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Solution

Backward Induction Algorithm:

Epoch 4: Set $t = T + 1 = 4$, $v_4^*(s) = 0 \; for \; s = 0,1,2,3$

Epoch 3: Since $t \neq 1 \;\; continue. Set \; t = 3.$

$$v_3^*(s) = \max_{a \in A}\{r(s,a) + \sum_{j \in S} p(j|s,a) v_4^*(j)\}, s = 0,1,2,3$$

$$= \max_{a \in A}\{\mathrm{r(s,a)}\}$$

It follows that $v_3^*(0) = 0, v_3^*(1) = 5, v_3^*(2) = 6, v_3^*(3) = 5$

and $a_3^*(s) = 0, for \; s = 0,1,2,3$

Solution Cont'd

Epoch 2: Since $t \neq 1$ $continue. Set\ t = 2.$

$$v_2^*(s) = \max_{a \in A}\{r(s,a) + \sum_{j \in S} p(j|s,a)v_3^*(j)\}, s = 0,1,2,3$$

$$= \max_{a \in A}\{v_2^*(s,a)\}$$

Since $v_2^*(s,a) =$

$$\begin{bmatrix} 0 & 0.25 & 2 & 0.5 \\ 6.25 & 4 & 2.5 & X \\ 10 & 4.5 & X & X \\ 10.5 & X & X & X \end{bmatrix}$$

it follows that $v_2^*(0) = 2, v_2^*(1) = 6.25, v_2^*(2) = 10, v_2^*(3) = 10.5\ and\ a_2^*(0) = 2, a_2^*(1) = 0, a_2^*(2) = 0, a_2^*(3) = 0$.

Solution Cont'd

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

Epoch 1: Since $t \neq 1$ $continue. Set\ t = 1.$

$$v_1^*(s) = \max_{a \in A}\{v_1^*(s,a)\}$$

$$v_1^*(s,a) = \begin{bmatrix} 2.00 & 2.06 & 4.12 & 4.19 \\ 8.06 & 6.12 & 6.19 & X \\ 12.12 & 8.19 & X & X \\ 14.19 & X & X & X \end{bmatrix}$$

It follows that $v_1^*(0) = 4.19, v_1^*(1) = 8.06, v_1^*(2) = 12.12,\ v_2^*(3) = 14.19\ and\ a_1^*(0) = 3, a_1^*(1) = 0, a_1^*(2) = 0, a_1^*(3) = 0$

Optimal Policy:

s	$d_1^*(s)$	$d_2^*(s)$	$d_3^*(s)$
0	3	2	0
1	0	0	0
2	0	0	0
3	0	0	0

If initial inventory is 0, order 3 units in period 1 and nothing after that.

Example

Consider a game where at the start of each of *n* turns you can bet some number of coins that you currently have. You will win with probability p and lose with probability 1-p. Your goal is to end the game with maximum expected reward.

1) Write the optimality equations in general form

2) Solve the problem for n=3, p = 0.75, and given you start with 3 coins

3) How do you think the results will change as p changes?