Stochastic Processes

Lecture 6: DTMC value function analysis

Hailun Zhang@SDS of CUHK-Shenzhen

January 27, 2021

Expected profit on day 3

- Expected profit on "day 3": $g(1) = -\$5$, $g(2) = \$1$, $g(3) = \$10$.

$$\begin{pmatrix} 0.4 & 0.48 & 0.12 \\ 0.3 & 0.4 & 0.3 \\ 0.6 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} -5 \\ 1 \\ 10 \end{pmatrix} = \begin{pmatrix} -0.32 \\ 1.90 \\ 1.00 \end{pmatrix}$$

Expected profit on day 3

- Expected profit on "day 3": $g(1) = -\$5$, $g(2) = \$1$, $g(3) = \$10$.

$$\begin{pmatrix} 0.4 & 0.48 & 0.12 \\ 0.3 & 0.4 & 0.3 \\ 0.6 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} -5 \\ 1 \\ 10 \end{pmatrix} = \begin{pmatrix} -0.32 \\ 1.90 \\ 1.00 \end{pmatrix}$$

- Thus,

$$\begin{pmatrix} \mathbb{E}[g(X_3)|X_0 = 1)] \\ \mathbb{E}[g(X_3)|X_0 = 2)] \\ \mathbb{E}[g(X_3)|X_0 = 3)] \end{pmatrix} = P^3 \begin{pmatrix} g(1) \\ g(2) \\ g(3) \end{pmatrix}$$

Expected total profit in three days

- $\mathbb{E}\Big[g(X_1) + g(X_2) + g(X_3) | X_0 = 1\Big]$

- $v^3(i) = \mathbb{E}\Big[g(X_1) + g(X_2) + g(X_3) | X_0 = i\Big]$, $i = 1, 2, 3.$

$$\begin{pmatrix} v^3(1) \\ v^3(2) \\ v^3(3) \end{pmatrix} = (P + P^2 + P^3) \begin{pmatrix} -5 \\ 1 \\ 10 \end{pmatrix} = \begin{pmatrix} 3.48 \\ 3.30 \\ -1.20 \end{pmatrix}$$

- Value function

Expected total discounted profit in three days

- Expected total discounted profit in three days

$$\mathbb{E}[\beta g(X_1) + \beta^2 g(X_2) + \beta^3 g(X_3) | X_0 = 1)]$$

100 Yuan on day 1 is worthy 95 Yuan today if $\beta = .95$.

- $v^3(i) = \mathbb{E}\Big[\beta g(X_1) + \beta^2 g(X_2) + \beta^3 g(X_3) | X_0 = i\Big]$, $i = 1, 2, 3$.

Non-perishable Inventory models

- Assume i.i.d. demand

d	0	1	2	3
$\mathbb{P}(D = d)$	1/8	1/4	1/2	1/8

- Assume our inventory policy to be $(s, S) = (2, 4)$.
- Let X_n be inventory level at the end of week n. Note that values that X_n can take is in $\{0, 1, 2, 3, 4\}$.
- $c_v = .25$, $c_f = .50$, $c_p = 1$, $c_s = -.10$.
- Find the expected profit in three weeks given $X_0 = 0$.

Expected total profit in multiple periods

- $g(i)$ is the expected profit in the next week, given the current week ends with i items.

$$g(0) = 11.1/8 - 1.5 = -0.1125$$

$$g(1) = 11.1/8 - 1.25 = 0.1375$$

$$g(2) = 2\left(\frac{4}{8} + \frac{1}{8}\right) + (1 - .1)\frac{2}{8} + (-.2)\frac{1}{8} = \frac{11.6}{8} = 1.45$$

$$g(3) = 3\frac{1}{8} + (2 - .1)\frac{4}{8} + (1 - .2)\frac{2}{8} + (-.3)\frac{1}{8} = 11.9/8 = 1.4875$$

$$g(4) = \quad = 1.3875$$

Expected profit in each period

- The expected profit in three periods given $X_0 = 0$ is

$$\mathbb{E}\Big[g(X_0) + g(X_1) + g(X_2) | X_0 = 0\Big]$$

- Recall

$$\begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{5}{8} & \frac{1}{4} & \frac{1}{8} & 0 & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & 0 \\ 0 & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \end{pmatrix} \end{array}$$

- $\mathbb{E}[g(X_1) | X_0 = 0]$
- $\mathbb{E}[g(X_2) | X_0 = 0]$

Why value function is important?

- AlphaGo Zero

$$v(i) = \mathbb{E}\Big[g(X_1) + g(X_2) + g(X_3) + \dots g(X_{\tau-1}) + h(X_\tau))|X_0 = i\Big]$$

- For airline yield management

$$v^{(10)}((10, 90))$$

expected profit if there are 10 days left for sale, 10 empty first class seats and 90 economic class seats.

Computation of the value function

- Suppose that $g(1) = -\$5$, $g(2) = \$1$, $g(3) = \$10$.

- $\mathbb{E}\Big[g(X_1) + g(X_2) + g(X_3) | X_0 = 1\Big]$

$$v(i) = \mathbb{E}\Big[g(X_0) + g(X_1) + g(X_2) + g(X_3) | X_0 = i\Big], \quad i = 1, 2, 3.$$

- value function

$$\begin{pmatrix} v(1) \\ v(2) \\ v(3) \end{pmatrix} = (I + P + P^2 + P^3) \begin{pmatrix} -5 \\ 1 \\ 10 \end{pmatrix} = \begin{pmatrix} -1.52 \\ 4.30 \\ 8.80 \end{pmatrix}$$

Complexity: a naive algorithm

- Let $m = |S|$ be the number of states.
- Computing Pg:
  for $i = 1$ to m
  $$(Pg)(i) = \sum_{\ell=1}^{m} P_{i\ell} g(\ell),$$
  requiring m^2 operations.
- Computing P^2 requires m^3 operations.
- Knowing P and P^2, computing P^3 requires m^3 operations.
- Knowing P, P^2, and P^3, computing $P + P^2 + P^3$ requires $2m^2$ operations.
- Knowing $P + P^2 + P^3$, computing $(P + P^2 + P^3)g$ requires m^2 operations.
- One algorithm to compute $(P + P^2 + P^3)g$ requires
  $$2m^3 + 3m^2$$
  operations. Think of $m = 1$ billion.

Complexity: Can we do better?

- Cost-to-go function v^k: $k = 0, 1, 2, 3$

$$v^k(j) = \mathbb{E}\Big[g(X_k) + \dots g(X_3) | X_k = j\Big], \quad j \in S.$$

- Then $v = v^0$.
- Algorithm

$$\begin{aligned} v^0 &= g + Pv^1, \\ v^1 &= g + Pv^2, \\ v^2 &= g + Pv^3, \quad m^2 + m \text{ operations} \\ v^3 &= g \end{aligned}$$

Bellman equation

Theorem

$$v^k = g + Pv^{k+1}, \quad k = 0, 1, 2$$

Proof.

Dynamic programming algorithm

- Backward induction: computing v^3, v^2, v^1, v^0 in this order.
- Complexity: $3(m^2 + m)$ operations.

When the problem is infinite horizon...

Expected total discounted profit over an infinite horizon

$$
\begin{aligned}
v(i) &= \mathbb{E}\Big[g(X_0) + \gamma g(X_1) + \gamma^2 g(X_2) + \ldots | X_0 = i\Big] \\
&= \mathbb{E}\Big[\sum_{n=0}^{\infty} \gamma^n g(X_n) | X_0 = i\Big], \quad i \in \mathcal{S}.
\end{aligned}
$$

Theorem (Bellman equation)

Assume that g is bounded.
(a) The value function satisfies that

$$
v = g + \gamma P v. \tag{1}
$$

(b) Solution v to (1) exists and is unique.

The theorem holds even if $|S|$ is infinite.

Proof of the theorem (a)