Advanced Topics in OR

Lecture Notes 8

Semi-Markov Decision Processes

Zhao Xiaobo
Department of IE
Tsinghua University
Beijing 100084, China
Tel. 010-62784898
Email. xbzhao@tsinghua.edu.cn

Introduction

If the process is in state i and action a is chosen, then

(i) The transition probability $P_{ij}(a)$

(ii) The time from i to j, a random variable with probability distribution $F_{ij}(\cdot|a)$

Immediate cost: $C(i, a)$, bounded

Cost rate: $c(i, a)$, bounded

Total cost: transition time t, $C(i, a) + t\, c(i, a)$

Condition 1: Avoid infinite number of transitions in finite interval

There exist $\delta > 0$, $\varepsilon > 0$, such that

$$\sum_{j=0}^{\infty} P_{ij}(a) F_{ij}(\delta|a) \le 1 - \varepsilon$$

Discounted Cost Criterion

α: discount rate

Cost C incurred at time $t \to Ce^{-\alpha t}$ at time 0

τ_n: time between the $(n-1)$st and the nth transition

$$V_\pi(i) = E_\pi\left[\sum_{n=1}^{\infty} e^{-\alpha(\tau_1+\cdots\tau_{n-1})}\left(C(X_n,a_n)+\int_0^{\tau_n} c(X_n,a_n)e^{-\alpha t}dt\right)\middle| X_1 = i\right]$$

Letting $\quad V_\alpha(i) = \inf_\pi V_\pi(i)$

π^* is α – optimal if $\; V_{\pi^*}(i) = V_\alpha(i), \quad$ for all i

Theorem 7.1

$$V_\alpha(i) = \min_a\left\{\bar{C}_\alpha(i,a) + \sum_{j=0}^{\infty} P_{ij}(a)\int_0^{\infty} e^{-\alpha t} V_\alpha(j)\, dF_{ij}(t|a)\right\}$$

Discounted Cost Criterion

where

$$\bar{C}_\alpha(i,a) = C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a) \int_0^\infty \int_0^t e^{-\alpha s} c(i,a)\, ds\, dF_{ij}(t|a)$$

is the expected one stage discounted cost

f_α: stationary policy, minimizing right side of $V_\alpha(i)$

Theorem 7.2

The stationary policy f_α is α – optimal. That is

$$V_{f_\alpha}(i) = V_\alpha(i), \quad \text{for all } i$$

Remark: For each stationary policy f, define the mapping

T_f: $B(I) \rightarrow B(I)$ by

Discounted Cost Criterion

$$\left(T_f u\right)(i) = \bar{C}_\alpha\left[i, f(i)\right] + \sum_{j=0}^{\infty} P_{ij}\left[f(i)\right]\int_0^\infty e^{-\alpha t} u(j)\, dF_{ij}\left[t \middle| f(i)\right]$$

Then by making use of Condition 1, the equivalent of Lemma 6.2 may be proven and then to prove Theorem 7.2

Define the mapping T_α: $B(I) \rightarrow B(I)$ by

$$\left(T_\alpha u\right)(i) = \min_a \left\{ \bar{C}_\alpha(i, a) + \sum_{j=0}^{\infty} P_{ij}(a) \int_0^\infty e^{-\alpha t} u(j)\, dF_{ij}\left(t \middle| a\right) \right\}$$

Theorem 7.3

$\|T_\alpha u - T_\alpha v\| \leq (1 - \varepsilon + \varepsilon e^{-\alpha\delta})\, \|u - v\|$ for all $u, v \in B(I)$.

$\rightarrow T_\alpha$ is a contraction mapping with fixed point V_α.

Average Cost - Preliminaries

$Z(t)$: total cost by time t

$Z_n = C(X_n, a_n) + \tau_n c(X_n, a_n)$: cost during the nth trnasition

For any policy π

$$\phi_\pi^1(i) = \lim_{t\to\infty} E_\pi\left[\frac{Z(t)}{t} \middle| X_1 = i\right]$$

and

$$\phi_\pi^2(i) = \lim_{t\to\infty} \frac{E_\pi\left[\sum_{j=1}^{n} Z_j \middle| X_1 = i\right]}{E_\pi\left[\sum_{j=1}^{n} \tau_j \middle| X_1 = i\right]}$$

φ^1 is the usual mean of average expected cost

φ^2 easier to work

Average Cost - Preliminaries

Is φ^1 equivalent to φ^2 ?

Under certain condition, they are.

Sufficient condition: For any stationary policy f, the resultant semi-Markov process $\{X(t), t \geq 0\}$ is a regenerative process with finite expected cycle length

Let $T = \min\left\{t > 0 : X(t) = i, X\left(t^{-}\right) \neq i\right\}$

$$N = \min\{n > 0 : X_{n+1} = i\}$$

Lemma 7.4 If $E_\pi[T|X_1 = i] < \infty$, then

$$E_\pi\left[N \middle| X_1 = i\right] < \infty, \quad \text{and} \quad T = \sum_{n=1}^{N} \tau_n$$

Average Cost - Preliminaries

Proof. It follows that $T \geq \sum_{n=1}^{N} \tau_n$ with equality if $N < \infty$

Let

$$\bar{\tau}_n = \begin{cases} 0 & \text{if } \tau_n \leq \delta \\ \delta \text{ with probability } \dfrac{\varepsilon}{1-\sum_{j=0}^{\infty} P_{kj}(a) F_{kj}(\delta|a)} & \text{if } \tau_n > \delta, X_n = k, a_n = a \\ 0 \text{ with probability } 1-\dfrac{\varepsilon}{1-\sum_{j=0}^{\infty} P_{kj}(a) F_{kj}(\delta|a)} & \text{if } \tau_n > \delta, X_n = k, a_n = a \end{cases}$$

From condition 1, $\bar{\tau}_n$ are iid with

$$P\{\bar{\tau}_n = \delta\} = \varepsilon = 1 - P\{\bar{\tau}_n = 0\}$$

Average Cost - Preliminaries

From Wald's equation

if $EN = \infty$ then $\quad E\sum_{n=1}^{N}\bar{\tau}_n = \infty$

$$\Longrightarrow \quad ET \geq E\sum_{n=1}^{N}\tau_n \geq E\sum_{n=1}^{N}\bar{\tau}_n = \infty$$

Therefore, if $ET < \infty$, then EN and hence N are finite

Theorem 7.5 If f is a stationary policy, and if $E_f[T|X_1 = i] < \infty$

$$\Longrightarrow \quad \phi_f^1(i) = \phi_f^2(i) = \frac{E_f\left[Z(T)\middle|X_1 = i\right]}{E_f\left[T\middle|X_1 = i\right]}$$

Proof. Under a stationary policy, $\{X(t), t > 0\}$ is a regenerative process with regeneration point T

Average Cost - Preliminaries

$\{Z(t), t > 0\}$: renewal reward process

$$\Rightarrow \quad \phi_f^1(i) = \lim_{t\to\infty} \frac{E_f Z(t)}{t} = \frac{E_f Z(T)}{E_f T}$$

$\{X_n, n = 1, 2, \ldots\}$: discrete time regenerative process with regeneration time N

$Z_1 + \cdots + Z_N$: reward during the first cycle

$$E_f \sum_{i=1}^{n} \frac{Z_i}{n} \to \frac{E_f \sum_{i=1}^{N} Z_i}{E_f N} \quad \text{as} \quad n \to \infty$$

Regard $\tau_1 + \cdots + \tau_N$ as reward during the first cycle

Average Cost - Preliminaries

$$E_f \sum_{i=1}^{n} \frac{\tau_i}{n} \to \frac{E_f \sum_{i=1}^{N} \tau_i}{E_f N} \quad \text{as} \quad n \to \infty$$

We obtain

$$\phi_f^2(i) = \frac{E_f \sum_{i=1}^{N} Z_i}{E_f \sum_{i=1}^{N} \tau_i}$$

However, since $N < \infty$, it is easy to see

$$\sum_{i=1}^{N} Z_i = Z(T) \qquad \sum_{i=1}^{N} \tau_i = T$$

the result follows

Average Cost - Preliminaries

Remarks: It follows, with probability 1

$$\lim_{t\to\infty}\frac{Z(t)}{t}=\lim_{n\to\infty}\frac{\sum_{i=1}^{n}Z_i}{\sum_{i=1}^{n}\tau_i}=\frac{E_f Z(T)}{E_f T}$$

When is it true that $\phi_f^1(j)=\phi_f^2(j)=\phi_f^1(i)$?

⇨ With probability 1, the process will eventually enter state i, then $\{X(t),\ t>0\}$ is a delayed regenerative process.

Additional notation $\bar{\tau}(i,a)=\sum_{j=0}^{\infty}P_{ij}(a)\int_0^{\infty}t\,dF_{ij}(t|a)$

$$\bar{C}(i,a)=C(i,a)+c(i,a)\bar{\tau}(i,a)$$

Average Cost - Preliminaries

$\bar{\tau}(i,a)$: the expected time until a transition occurs

$\bar{C}(i,a)$: the expected cost during such a transition

φ^2 only depends on the parameters of the process through the three functions $\bar{\tau}(i,a)$, $\bar{C}(i,a)$, $P_{ij}(a)$

⇨ Choose cost and transition time distributions in as convenient a manner as possible

Without loss of generality, assume

$$C(i,a) = \bar{C}(i,a) \qquad c(i,a) = 0$$

and the time until transition is (with probability 1) $\bar{\tau}(i,a)$

Average Cost - Results

Theorem 7.6 If there exists a bounded function $h(i)$ and a constant g such that

$$h(i)=\min_{a}\left\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)h(j)-g\bar{\tau}(i,a)\right\}$$

then there exists a stationary π^* such that

$$g=\phi_{\pi^*}^{2}(i)=\min_{\pi}\phi_{\pi}^{2}(i) \quad \text{for all } i$$

Proof. Let $H_n=(X_1,a_1,\cdots,X_n,a_n)$

For any policy π, $$E_{\pi}\left\{\sum_{i=2}^{n}\left[h(X_i)-E_{\pi}\left(h(X_i)\middle|H_{i-1}\right)\right]\right\}=0$$

Average Cost - Results

But
$$E_\pi\left[h(X_i)\middle|H_{i-1}\right]=\sum_{j=0}^{\infty}h(j)P_{X_{i-1}j}(a_{i-1})$$
$$=\bar{C}(X_{i-1},a_{i-1})+\sum_{j=0}^{\infty}h(j)P_{X_{i-1}j}(a_{i-1})-g\bar{\tau}(X_{i-1},a_{i-1})$$
$$-\bar{C}(X_{i-1},a_{i-1})+g\bar{\tau}(X_{i-1},a_{i-1})$$
$$\geq\min_a\left\{\bar{C}(X_{i-1},a)+\sum_{j=0}^{\infty}h(j)P_{X_{i-1}j}(a)-g\bar{\tau}(X_{i-1},a)\right\}$$
$$-\bar{C}(X_{i-1},a_{i-1})+g\bar{\tau}(X_{i-1},a_{i-1})$$
$$=h(X_{i-1})-\bar{C}(X_{i-1},a_{i-1})+g\bar{\tau}(X_{i-1},a_{i-1})$$

Average Cost - Results

with equality for π^*. Hnece

$$0 \le E_\pi \left\{ \sum_{i=2}^{n} \left[h(X_i) - h(X_{i-1}) + C(X_{i-1}, a_{i-1}) - g\bar{\tau}(X_{i-1}, a_{i-1}) \right] \right\}$$

$$\Rightarrow \quad g \le \frac{E_\pi \left[h(X_n) - h(X_1) \right] + E_\pi \sum_{i=2}^{n} \bar{C}(X_{i-1}, a_{i-1})}{E_\pi \sum_{i=2}^{n} \bar{\tau}(X_{i-1}, a_{i-1})}$$

with equality for π^*.

Letting $n \to \infty$ and using the boundedness of h and the fact that Condition 1 implies

$$E_\pi \sum_{i=2}^{n} \bar{\tau}(X_{i-1}, a_{i-1}) \ge (n-1)\varepsilon\delta \to \infty$$

Average Cost - Results

we have

$$g \le \lim_{n\to\infty} \frac{E_\pi \sum_{i=2}^{n} \bar{C}\left(X_{i-1}, a_{i-1}\right)}{E_\pi \sum_{i=2}^{n} \bar{\tau}\left(X_{i-1}, a_{i-1}\right)} = \phi_\pi^2\left(X_i\right)$$

with equality for π^* and all values of X_1.

When the conditions of Theorem 6.7 are satisfied?

We have assumed that (without loss of generality)

$C(i,a) = \bar{C}(i,a)$ $\quad c(i,a) = 0$ $\quad$ transition time $\bar{\tau}(i,a)$

$$\Rightarrow \quad V_\alpha(i) = \min_a \left\{ \bar{C}(i,a) + e^{-\alpha\bar{\tau}(i,a)} \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\}$$

Average Cost - Results

Fix state 0, and define $h_\alpha(i) = V_\alpha(i) - V_\alpha(0)$

Then, we obtain

$$h_\alpha(i) = \min_a \left\{ \bar{C}(i,a) + e^{-\alpha\bar{\tau}(i,a)} \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) + \left[e^{-\alpha\bar{\tau}(i,a)} - 1 \right] V_\alpha(0) \right\}$$

$$= \min_a \left\{ \bar{C}(i,a) + e^{-\alpha\bar{\tau}(i,a)} \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) - V_\alpha(0) \left[\alpha\bar{\tau}(i,a) + o(\alpha) \right] \right\}$$

Theorem 7.7 If $|V_\alpha(i) - V_\alpha(0)| < N$ for all α, all i

(i) Exist bounded $h(i)$ and constant g satisfying (6)

(ii) For $\alpha_n \to 0$, $h(i) = \lim_{n\to\infty} \left(V_{\alpha_n}(i) - V_{\alpha_n}(0) \right)$

(iii) $\lim_{\alpha\to 0} \alpha V_{\alpha_n}(0) = g$

Some Examples

Letters arrive at post office ~ Poisson process with rate λ

Action: (i) summon a truck to pick up all letters, cost K

(ii) wait, cost rate $C(i)$ bounded increasing nonnegative

Problem: select a policy, minimize the long-run average cost

SMDP: state i, the number of letters in the post

action 1: summon a truck

action 2: don't summon a truck

$$P_{i1}(1)=1 \qquad \bar{\tau}(i,1)=1/\lambda \qquad \bar{C}(i,1)=K+C(0)/\lambda$$

$$P_{ii+1}(2)=1 \qquad \bar{\tau}(i,2)=1/\lambda \qquad \bar{C}(i,2)=C(i)/\lambda$$

Some Examples

Let
$$V_\alpha(i,1)=\min\left\{K+\frac{C(0)}{\lambda};\frac{C(i)}{\lambda}\right\}$$

and for $n > 1$

$$V_\alpha(i,n)=\min\left\{K+\frac{C(0)}{\lambda}+e^{-\alpha/\lambda}V_\alpha(1,n-1);\frac{C(i)}{\lambda}+e^{-\alpha/\lambda}V_\alpha(i+1,n-1)\right\}$$

By induction method, $V_\alpha(i, n)$ is increasing in i

$\Rightarrow$ $V_\alpha(i) = \lim V_\alpha(i, n)$ is increasing in i

Since $V_\alpha(i)$ satisfies

$$V_\alpha(i)=\min\left\{K+\frac{C(0)}{\lambda}+e^{-\alpha/\lambda}V_\alpha(1);\frac{C(i)}{\lambda}+e^{-\alpha/\lambda}V_\alpha(i+1)\right\}$$

Some Examples

It follows that $$V_\alpha(i) \le K + \frac{C(0)}{\lambda} + e^{-\alpha/\lambda} V_\alpha(1)$$

$$< K + \frac{C(0)}{\lambda} + V_\alpha(1)$$

$$\Rightarrow \quad V_\alpha(1) < V_\alpha(i) < K + \frac{C(0)}{\lambda} + V_\alpha(1)$$

From Theorem 7.7, there exist a constant g and bounded increasing function $h(i)$

$$h(i) = \min\left\{K + \frac{C(0)}{\lambda} + h(1) - \frac{g}{\lambda}; \frac{C(i)}{\lambda} + h(i+1) - \frac{g}{\lambda}\right\}$$

Some Examples

Let
$$i^* = \min\left\{ i : \frac{C(i)}{\lambda} + h(i+1) > K + \frac{C(0)}{\lambda} + h(1) \right\}$$

From monotonicity of $C(i)$ and $h(i)$, $\rightarrow$ summon a truck whenever the number of letters in the post is at lest i^*

Determine i^*

f_i: policy, summon a truck whenever at least i letters

⇨ Regenerative process, state 1

The long-run average cost

$$\phi_{f_i}(j) = \frac{E_{f_i}[\text{cost of cycle}]}{E_{f_i}[\text{length of cycle}]} = \frac{K + \frac{C(0)}{\lambda} + E\int_{\tau_1}^{\tau_i} C[N(t)]dt}{\frac{i}{\lambda}}$$

Some Examples

Hence

$$\phi_{f_i}(j) = \frac{K + \dfrac{C(0)}{\lambda} + E\left[C(1)(\tau_2 - \tau_1) + \cdots + C(i-1)(\tau_i - \tau_{i-1})\right]}{\dfrac{i}{\lambda}}$$

$$= \frac{\lambda}{i}\left[K + \frac{C(0)}{\lambda} + \sum_{j=1}^{i-1} \frac{C(j)}{\lambda}\right]$$

$$= \frac{\lambda K}{i} + \frac{1}{i}\sum_{j=0}^{i-1} C(j)$$

As an example, if $C(i) = iC$, then $\phi_{f_i}(j) = \dfrac{\lambda K}{i} + \dfrac{(i-1)C}{2}$

The optimal i is one of the two integers adjacent to $\sqrt{2\lambda K/C}$

Some Examples

The streetwalker' dilemma

Customers arrive ~ Poisson process with rate λ

offer pair (i, F_i): i the money

F_i distribution of service time with offer i

$$t_i = \int_0^\infty x dF_i(x)$$

(i, F_i) occurs with probability P_i

SMDP: state i,

action 1: accept

action 2: reject

Some Examples

The streetwalker' dilemma

Customers arrive ~ Poisson process with rate λ

$$P_{ij}(1) = P_j \qquad \bar{\tau}(i,1) = t_i + 1/\lambda \qquad \bar{C}(i,1) = -i$$

$$P_{ij}(2) = P_j \qquad \bar{\tau}(i,2) = 1/\lambda \qquad \bar{C}(i,2) = 0$$

It is easy to check the conditions of Theorem 7.7 are satisfied, hence by Theorem 7.6, we have

$$h(i) = \min\left\{ -i + \sum_{j=1}^{N} P_j h(j) - g\left(t_i + \frac{1}{\lambda}\right); \sum_{j=1}^{N} P_j h(j) - g\frac{1}{\lambda} \right\}$$

The optimal policy accepts an offer (i, F_i) iff $\dfrac{i}{t_i} \geq g$