سوال ۱: با رابطه زیر داریم:

$$V^{\pi}(s) = \sum_{s'} T(s,\pi(s),s')\left[R(s,\pi(s),s') + \gamma V^{\pi}(s')\right]$$

$$\pi_{i+1}(s) = \arg\max_{a} \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^{\pi_i}(s')\right]$$

در ابتدا $V(s)$ همه برابر صفر هستند.

	s_1	s_2	s_3	s_4	s_5	s_6
$V_0(s)$	0	0	0	0	0	0

حالت پایه این شد π_0 برای همه حالت‌هایشان finish باشد.

	s_1	s_2	s_3	s_4	s_5	s_6
$V_0(s)$	0	0	0	0	0	0
π_0	finish	finish	finish	finish	finish	finish
$V_1^{\pi_0}(s)$	1	2	3	4	5	6
π_1	dice	dice	finish	finish	finish	finish
$V_2^{\pi_1}(s)$	2.5	2.51	3	4	5	6

در صفحه بعد ادامه داده شده

$$\pi_1(s_1) = \arg\max_{a} \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^{\pi_0}(s')\right] =$$

$$\max\left(\underset{\text{finish}}{1}, \frac{1}{6}\left(\cancel{6} + \underbrace{1+2+3+4+5+\cancel{6}}\right)\right) = \text{dice}$$

$$\frac{15}{6}$$

$$\pi_1(s_2) = \left(2, \frac{15}{6}\right) = \text{dic}$$

اما در s_3 به بعد همان حالت finish را دارند.

چون $3 > \frac{15}{6}$ هست.

ادامه سوال ①

$$v_2^{\pi_1(s_1)} = \left(\frac{1}{6}(-1+1) + \frac{1}{6}(-1+2) + \dots + \frac{1}{6}(-1+6) \right)$$

$$= \left(0 + \frac{1}{6} + \frac{2}{6} + \dots + \frac{5}{6} \right) = \frac{15}{6} = \boxed{2.5}$$

$$v_2^{\pi_1(s_2)} = \;\; \longrightarrow$$

$\boxed{v_2^{\pi_1(s_1)}}$ مثل

← حالا π_2 را حساب می‌کنیم

$$\pi_2(s_1) = \max\left(\underset{\text{finish}}{2.5},\ \underset{\text{dice}}{\frac{1}{6}\left(-6 + 2.\cancel{5}^{5} + \cancel{2.5} + 3 + 4 + 5 + 6\right)} \right) = \underline{\text{dice}}$$

$\frac{17}{6}$

باز برای $\pi_2(s_1)$ ، $\pi_2(s_2)$ برابر با $\underline{\text{dice}}$ می‌شود و برای بقیه همان finish است.

	s_1	s_2	s_3	s_4	s_5	s_6
π_2	dice	dice	finish	finish	finish	finish
$v_3^{\pi_2(s)}$	2.833	2.833	3	4	5	6

$$v_3^{\pi_2(s_1)} = \left(\frac{1}{6}(-6 + 2.5 + 2.5 + 3 + 4 + 5 + 6) \right) = \frac{17}{6} = 2.833$$

حالا برای $\underline{\pi_3}$ حساب می‌کنیم

$$\pi_3(s_1) = \max\left(\underset{\text{fisih}}{2.833},\ \overset{\text{dice}}{\frac{1}{6}(-6 + 2.833 \times 2 + 18)} \right) =$$

2.94

	s_1	s_2	s_3	s_4	s_5	s_6
π_3	dice	dice	finish	finish	finish	finish
$v_4^{\pi_3(s)}$	2.94	2.94	3	4	5	6

$$\pi_4 : \left(2.94,\ \frac{1}{6}\left(-6 + \overset{\text{dice}}{2\times 2.94 + 18}\right)\right) \Rightarrow \underline{\text{dice}}$$

$\underset{\text{finish}}{2.94}$ ، 2.98

	s_1	s_2	s_3	s_4	s_5	s_6
π_4	dice	dice	finish	finish	finish	finish
$v_5^{\pi_4(s)}$	2.98	2.98	3	4	5	6

اگر دقت بیشتر شود مقدار s_1, s_2 نزدیک به $\underline{3}$ میل می‌کنند. زیرا از این قرار تقریبا همگرا شدیم. پس جدول به این صورت

بیشتر نقطه یکسان صورت می‌گیرد

چون s_1, s_2 به $\underline{3}$ میل می‌کنند $\pi_n(s_3)$ را به تو هم $\underline{\text{finish}}$ باشد هم dice

$$\max\left(3,\ \frac{1}{6}\left(-6 + 2\times 3 + 18\right)\right)$$

هر دو یکی می‌شه

صورت	s_1	s_2	s_3	s_4	s_5	s_6
π_2	dice	dice	finish	finish	finish	finish
v^{π_2}	با تقریب $2.98 \approx 3$	$2.98 \approx 3$	3	4	5	6

قسمت دوم:

صورت	s_1	s_2	s_3	s_4	s_5	s_6
π_2	dice	dice	finish	finish	finish	finish
π_{i+1}	dice	dice	dice/finish	finish	finish	finish

قسمت سوم: از آنجایی که پالیسی π_2 نسبت به π_{i+1} تغییر نکرده (در $\pi_{i+1}(s_3)$ هم هر دو پالیسی درسته) پس پالیسی‌ها $\underline{\text{converge}}$ شده است (اگر دقت بیشتر در $\underline{\pi_1}$ برای پالیسی‌ها رسیده بودیم با v^{π_2} ها است هنوز تغییر می‌کرد $v^{\pi(s_1)}$ ، $v^{\pi_i(s_2)}$ ها است که به $\underline{3}$ همگرا شوند.

سوال ۲

قسمت ۱ : Policy بهینه این هست که از state 1 تا state n-1 به سمت راست برود هر وقت به state n-1 رسید reset شود و به state1 باز گردد. این چرخه را هی تکرار کند. در این حالت بیشترین utility را به دست می‌آوریم. در واقع همان استارت

قسمت ۲

$$v^*(n) = \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma v^*(s') \right]$$

در state n فقط می‌توانیم یک action را انجام دهیم. اسم آن را exit بگذاریم و برمی‌گردیم یا خارج می‌شویم.

$$v^*(n) = \left(1 \left[10 + \frac{1}{2} \times v^*(n) \right] \right) \Rightarrow v^*(n) = 10 + \frac{v^*(n)}{2}$$

$$\rightarrow \frac{v^*(n)}{2} = 10 \rightarrow \boxed{v^*(n) = 20}$$

قسمت ۳ :

$$v^*(1) = \max \left(\overset{\text{reset}}{\frac{1}{2} \left(0 + \frac{1}{2} v^*(1) \right)} , \overset{\text{right}}{\frac{1}{2} \left(1 + \frac{1}{2} v^*(2) \right)} \right)$$

$$v^*(1) = \max \left(\frac{v^*(1)}{4} , \frac{v^*(2)+2}{4} \right)$$

$$v^*(2) = \max \left(\frac{v^*(1)}{4} , \frac{v^*(3)+2}{4} \right)$$

$$\vdots$$

$$v^*(n-2) = \max \left(\frac{v^*(1)}{4} , \frac{v^*(n-1)+2}{4} \right)$$

$$v^*(n-1) = \max \left(\frac{v^*(1)}{4} , \frac{20+2}{4} = \frac{22}{4} \right)$$

ادامه در صفحه بعد.

ادامه قسمت 3 : اگر که $V^*(n-1) = \max\left(\frac{V^*(1)}{2}, \frac{22}{4}\right)$ برابر $\frac{V^*(1)}{2}$ بشود دیگر جست ماکسیمم بقیه

هم $\frac{V^*(1)}{2}$ می‌شود در $V^*(1) = \frac{V^*(1)}{2}$ می‌شود پس $V^*(1) = 0$ بنابراین همه state ها برابر با صفر می‌شود پس

$$\boxed{V^*(n-1) = \frac{22}{4}}$$

و در کل :

$$\boxed{V^*(k) = \frac{20 + 2\times\left(4^{n-k-1} + 1\right)}{4^{n-k}}}$$

قسمت 4 : در اینجا باید از فرمول

$$V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V_k(s')\right]$$

	s_1	s_2	...	s_{n-1}	s_n
$V_0(s)$	0	0	...	0	0
$V_1(s)$	$\frac{1}{2}$	$\frac{1}{2}$	...	$\frac{1}{2}$	5

$$V_1(s_1) \leftarrow \max\Big(\underset{\text{reset}}{\frac{1}{2}\left(0 + V_0(s_1)\right)}, \underset{\text{right}}{\frac{1}{2}(1 + 0)}\Big) = \frac{1}{2}$$

برای 2 تا n-1 هم به طریق مشابه عمل می‌کنیم و همین $\frac{1}{2}$ می‌شود

$$V_1(s_n) = 1\times(10 + 0) = 5 \rightarrow V_1(s_n) = 0$$

پس $V_1(s_1)$ تا $V_1(s_{n-1})$ برابر $\frac{1}{2}$ است و $V_1(s_n)$ برابر با 5 است

$$V_2(s_1) = \max\left(\frac{1}{2}\left(0 + 1\times\frac{1}{2}\right), \frac{1}{2}\left(1 + 1\times\frac{1}{2}\right)\right) = \boxed{\frac{5}{8}}$$

$\frac{1}{8}$, $\frac{5}{2}$

پس $V_2(s_1)$ تا $V_2(s_{n-1})$ برابر $\frac{5}{8}$ است

$$V_2(s_n) = 1\left(10 + \frac{5}{2}\right) = \boxed{\frac{25}{2}}$$

	s_1	...	s_{n-1}	s_n
$V_0(s)$	0	...	0	0
$V_1(s)$	$\frac{1}{2}$	...	$\frac{1}{2}$	5
$V_2(s)$	$\frac{5}{8}$	...	$\frac{5}{8}$	$\frac{25}{2}$

5