1 Q1. Pacman MDP

1.1 (a) Single Food Pellet Grid

Grid layout: 2×3 with food at F, discount $\gamma = 0.5$

A	B	C
D	E	F

(i) Optimal Policy (1 point)
Using Bellman equation: $V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$
Q-value analysis:

- At A: Q(A,East) = Q(A,South) = 0.125
- At B: Q(B,East) = Q(B,South) = 0.25

State	$\pi(state)$
A	East or South
B	East or South
C	South
D	East
E	East

(ii) Optimal Value V*(A) (1 point)
Working backwards from F:

$$\begin{aligned}
V^*(F) &= 0 \\
V^*(E) &= 1 \\
V^*(C) &= 1 \\
V^*(B) &= 0 + 0.5 \times 1 = 0.5 \\
V^*(D) &= 0 + 0.5 \times 1 = 0.5 \\
V^*(A) &= 0 + 0.5 \times 0.5 = 0.25
\end{aligned}$$

$\boxed{V^*(A) = 0.25}$

(iii) Value Iteration Convergence (1 point)

k	V(A)	V(B)	V(C)	V(D)	V(E)	V(F)
0	0	0	0	0	0	0
1	0	0	1	0	1	0
2	0	0.5	1	0.5	1	0
3	0.25	0.5	1	0.5	1	0
4	0.25	0.5	1	0.5	1	0

$\boxed{k = 3}$

1.2 (b) Cherries at D and F

State space: $(position, cherry\_D, cherry\_F)$ where cherry status is 0/1, for disappeared and present.

(i) Optimal Policy with $\gamma = 1$, living reward = -1 (1 point)

States with cherries worth 5 points, dot worth 1 point, living cost -1.

State	$\pi(state)$
(A,1,1)	South
(A,1,0)	South
(A,0,1)	East
(A,0,0)	East
(C,1,1)	East
(C,1,0)	East
(C,0,1)	East
(C,0,0)	North or East
(D,0,1)	East
(D,0,0)	North
(E,1,1)	East
(E,1,0)	West
(E,0,1)	East
(E,0,0)	West
(F,1,0)	West
(F,0,0)	West

(ii) Living Reward Range for Exactly One Cherry (1 point)

For A→C→D→B to be better than both:

- Better than A→B: $6 + 3r > 1 + r \Rightarrow r > -2.5$
- Better than both cherries: $6 + 3r > 11 + 7r \Rightarrow r < -1.25$

For exactly one cherry: $\boxed{-2.5 < r < -1.25}$

2 Q2. MDP with Exit Action

Grid layout:

A	B
D	C

2.1 (a) Deterministic Actions, $\gamma = \frac{1}{2}$ (1 point)

Using Bellman equation with deterministic transitions. $V_D = x$. Need to solve system:

Let V_A, V_B, V_C be the values. Then:

$$V_A = 1 + \frac{1}{2}\max(V_B, x)$$
$$V_B = 1 + \frac{1}{2}\max(V_A, V_C)$$
$$V_C = 1 + \frac{1}{2}\max(V_B, x)$$

Since V_A and V_C have same equation, $V_A = V_C = V$. Then: $V_B = 1 + \frac{V}{2}$ and $V = 1 + \frac{1}{2}\max(V_B, x) = 1 + \frac{1}{2}\max(1 + \frac{V}{2}, x)$

Case 1: $x \leq 2 \to V = 1 + \frac{1}{2}(1 + \frac{V}{2}) = \frac{3}{2} + \frac{V}{4} \to V = 2$

Case 2: $x > 2 \to V = 1 + \frac{x}{2}$ and $V_B = \frac{3}{2} + \frac{x}{4}$

$$\boxed{V^*(D) = x} \tag{1}$$

$$\boxed{V^*(A) = V^*(C) = \begin{cases} 2 & \text{if } x \leq 2 \\ 1 + \frac{x}{2} & \text{if } x > 2 \end{cases}} \tag{2}$$

$$\boxed{V^*(B) = \begin{cases} 2 & \text{if } x \leq 2 \\ \frac{3}{2} + \frac{x}{4} & \text{if } x > 2 \end{cases}} \tag{3}$$

2.2 (b) Stochastic Actions, Success Probability = $\frac{1}{2}$ (1 point)

With 50% success rate, otherwise stay with reward 0. Need new value functions for stochastic case:

Let V_A^s, V_B^s, V_C^s be stochastic values. Then:

$$V_A^s = \max(\frac{1}{2}[1 + \frac{x}{2}] + \frac{1}{4}V_A^s, \frac{1}{2}[1 + \frac{V_B^s}{2}] + \frac{1}{4}V_A^s)$$
$$V_B^s = \max(\frac{1}{2}[1 + \frac{V_A^s}{2}] + \frac{1}{4}V_B^s, \frac{1}{2}[1 + \frac{V_C^s}{2}] + \frac{1}{4}V_B^s)$$
$$V_C^s = \max(\frac{1}{2}[1 + \frac{x}{2}] + \frac{1}{4}V_C^s, \frac{1}{2}[1 + \frac{V_B^s}{2}] + \frac{1}{4}V_C^s)$$

For $Q^*(A; DOWN) = Q^*(A; RIGHT)$: $\frac{1}{2} + \frac{x}{4} + \frac{V_A^s}{4} = \frac{1}{2} + \frac{V_B^s}{4} + \frac{V_A^s}{4}$

This simplifies to: $x = V_B^s$

With $V_A^s = V_C^s = V$ and $V_B^s = x$:

From A: $V = \frac{1}{2} + \frac{x}{4} + \frac{V}{4} \Rightarrow \frac{3V}{4} = \frac{1}{2} + \frac{x}{4} \Rightarrow V = \frac{2}{3} + \frac{x}{3}$

From B: $x = \frac{1}{2} + \frac{V}{4} + \frac{x}{4} \Rightarrow \frac{3x}{4} = \frac{1}{2} + \frac{V}{4}$

Substituting: $\frac{3x}{4} = \frac{1}{2} + \frac{1}{4}(\frac{2}{3} + \frac{x}{3}) = \frac{1}{2} + \frac{1}{6} + \frac{x}{12} = \frac{2}{3} + \frac{x}{12}$

Solving: $\frac{3x}{4} - \frac{x}{12} = \frac{2}{3} \Rightarrow \frac{9x - x}{12} = \frac{2}{3} \Rightarrow \frac{8x}{12} = \frac{2}{3} \Rightarrow x = 1$

$\boxed{x = 1}$

2.3 (c) Modified Bellman Equation with Dice (1 point)

With dice roll adding 1-6 to reward (expected value = 3.5):

$$\boxed{V_{k+1}(s) = \max_a \sum_{s'} T(s, a, s')[R(s, a, s') + 3.5 + \gamma V_k(s')]}$$