



✓ Поздравляем! Вы прошли тест!  
для УСПЕШНОГО ПРОХОЖДЕНИЯ 80% или выше

Продолжить обучение

ОЦЕНКА  
100%

## Value Functions and Bellman Equations

ОБЩИЙ БАЛЛ 10

1. A policy is a function which maps \_\_\_\_ to \_\_\_\_.

1 / 1 балл

- ☐ States to values.
- ☐ Actions to probability distributions over values.
- ☒ States to probability distributions over actions.
- ☐ Actions to probabilities.
- ☐ States to actions.

✓ Correct  
Correct!

2. The term "backup" most closely resembles the term \_\_\_\_ in meaning.

1 / 1 балл

- ☐ Value
- ☒ Update
- ☐ Diagram

✓ Correct  
Correct!

3. At least one deterministic optimal policy exists in every Markov decision process.

1 / 1 балл

- ☒ True
- ☐ False

✓ Correct  
Correct! Let's say there is a policy  $\pi_1$  which does well in some states, while policy  $\pi_2$  does well in others. We could combine these policies into a third policy  $\pi_3$ , which always chooses actions according to whichever of policy  $\pi_1$  and  $\pi_2$  has the highest value in the current state.  $\pi_3$  will necessarily have a value greater than or equal to both  $\pi_1$  and  $\pi_2$  in every state! So we will never have a situation where doing well in one state requires sacrificing value in another. Because of this, there always exists some policy which is best in every state. This is of course only an informal argument, but there is in fact a rigorous proof showing that there must always exist at least one optimal deterministic policy.

4. The optimal state-value function:

1 / 1 балл

- ☐ Is not guaranteed to be unique, even in finite Markov decision processes.
- ☒ Is unique in every finite Markov decision process.

✓ Correct  
Correct! The Bellman optimality equation is actually a system of equations, one for each state, so if there are  $N$  states, then there are  $N$  equations in  $N$  unknowns. If the dynamics of the environment are known, then in principle one can solve this system of equations for the optimal value function using any one of a variety of methods for solving systems of nonlinear equations. All optimal policies share the same optimal state-value function.

5. Does adding a constant to all rewards change the set of optimal policies in episodic tasks?

1 / 1 балл

- ☒ Yes, adding a constant to all rewards changes the set of optimal policies.
- ☐ No, as long as the relative differences between rewards remain the same, the set of optimal policies is the same.

✓ Correct  
Correct! Adding a constant to the reward signal can make longer episodes more or less advantageous (depending on whether the constant is positive or negative).

6. Does adding a constant to all rewards change the set of optimal policies in continuing tasks?

1 / 1 балл

- ☒ No, as long as the relative differences between rewards remain the same, the set of optimal policies is the same.
- ☐ Yes, adding a constant to all rewards changes the set of optimal policies.

✓ Correct  
Correct! Since the task is continuing, the agent will accumulate the same amount of extra reward independent of its behavior.

reward independent of its behavior.

7. Select the equation that correctly relates  $v_\pi$  to  $q_\pi$ . Assume  $\pi$  is the uniform random policy.

1 / 1 Gann

- ☐  $v_\pi(s) = \sum_{a,s',r} \pi(as)p(s',rls,a)q_\pi(s')$
- ☐  $v_\pi(s) = \sum_{a,s',r} \pi(as)p(s',rls,a)[r + \gamma q_\pi(s')]$
- ☐  $v_\pi(s) = \sum_{a,s',r} \pi(as)p(s',rls,a)[r + q_\pi(s')]$
- ☒  $v_\pi(s) = \max_a q_\pi(s,a)$

✓ Correct  
Correct!

8. Select the equation that correctly relates  $q_\pi$  to  $v_\pi$  using four-argument function  $p$ .

1 / 1 Gann

- ☐  $q_\pi(s,a) = \sum_{s',r} p(s',rla,s)[r + v_\pi(s')]$
- ☐  $q_\pi(s,a) = \sum_{s',r} p(s',rla,s)\gamma[r + v_\pi(s')]$
- ☒  $q_\pi(s,a) = \sum_{s',r} p(s',rla,s)[r + \gamma v_\pi(s')]$

✓ Correct  
Correct!

9. Write a policy  $\pi_\pi$  in terms of  $q_\pi$ .

1 / 1 Gann

- ☐  $\pi_\pi(as) = q_\pi(s,a)$
- ☐  $\pi_\pi(as) = \max_{a'} q_\pi(s,a')$
- ☒  $\pi_\pi(as) = 1$  if  $a = \operatorname{argmax}_{a'} q_\pi(s,a')$ , else 0

✓ Correct  
Correct!

10. Give an equation for some  $\pi_\pi$  in terms of  $v_\pi$  and the four-argument  $p$ .

1 / 1 Gann

- ☐  $\pi_\pi(as) = 1$  if  $v_\pi(s) = \max_{a'} \sum_{s',r} p(s',rls,a')[r + \gamma v_\pi(s')]$ , else 0
- ☐  $\pi_\pi(as) = \sum_{s',r} p(s',rls,a)[r + \gamma v_\pi(s')]$
- ☐  $\pi_\pi(as) = \max_{a'} \sum_{s',r} p(s',rls,a')[r + \gamma v_\pi(s')]$
- ☒  $\pi_\pi(as) = 1$  if  $v_\pi(s) = \sum_{s',r} p(s',rls,a)[r + \gamma v_\pi(s')]$ , else 0

✓ Correct  
Correct!