Problem 1

Tickets to a lottery cost $1. There are two possible prizes: a $10 payoff with probabil-ity 1/50, and a $1,000,000 payoff with probability 1/2,000,000. What is the expected mone-tary value of a lottery ticket? When (if ever) is it rational to buy a ticket? Be preciseshow anequation involving utili-ties. You may assume current wealth of $kand thatU(Sk)=0.Youmay also assume thatU(Sk+10)=10U(Sk+1), but you may not make any assumption-saboutU(Sk+1,000,000). Sociological studies show that people with lower income buy a dis-proportionate number of lottery tickets. Do you think this is because they are worse decisionmakers or because they have a different util-ity function? Consider the value of contemplating the possibility of winning the lottery versus the value of contemplating becoming an actionhero while watching an adventure movie.

Answer words

Problem 2

In 1713, Nicolas Bernoulli stated a puzzle, now called the St. Petersburg paradox,which works as follows. You have the opportunity to play a game in which a fair coin is tossed repeatedly until it comes up heads. If the first heads appears on the nth toss, you win 2n dollars.

Part a **Show that the expected monetary value of this game is infinite.**

If n is infinite, then

$$2(\infty)0.5 = \infty$$

Part b **How much would you, personally, pay to play the game?**

I would pay $1, so if it lands heads first, I get my money back, after that I can only make a profit

Part c **Nicolass cousin Daniel Bernoulli resolved the apparent paradox in 1738 by suggestingthat the utility of money is measured on a logarithmic scale (i.e.,U(Sn)=alog2n+b,whereSnis the state of having $n). What is the expected utility of the game under this assumption?**

Answer words

Part d **What is the maximum amount that it would be rational to pay to play the game, assum-ing that ones initial wealth is $k**

Answer words

Problem 3

A used-car buyer can decide to carry out various tests with various costs (e.g., kick the tires, take the car to a qualified mechanic) and then, depending on the outcome of the tests, decide which car to buy. We will assume that the buyer is deciding whether to buy car c1, that there is time to carry out at most one test, and that t1 is the test of c1 and costs \$50.A car can be in good shape (quality q^+) or bad shape (quality q-), and the tests might help indicate what shape the car is in. Car c1 costs \$1,500, and its market value is \$2,000 if it is in good shape; if not, \$700 in repairs will be needed to make it in good shape. The buyers estimate is that c1 has a 70% chance of being in good shape

<u>Part a</u> **Calculate the expected net gain from buying c1, given no test.**

With no test, the expected value will be

$$(0.7)(2000 - 1500) + (0.3)(2000 - (1500 + 700) = \$290$$

<u>Part b</u> tests can be described by the probability that the car will pass or fail the test given thatthe car is in good or bad shape. We have the following information:

$$P(pass|q^+ = 0.8$$

$$P(pass|q^- = 0.35$$

Use Bayes theorem to calculate the probability that the car will pass (or fail) its test andhence the probability that it is in good (or bad) shape given each possible test outcome.

We need to calculate

$$P(q^+|pass) = \frac{(P(pass|q^+)(P(q^+)))}{P(pass)}$$

$$P(q^-|pass) = \frac{(P(pass|q^-(P(q^-)}{P(pass)}$$

And we know

$$P(q^+) = 0.7$$

$$P(q^-) = 0.3$$

$$P(pass|q^+) = 0.8$$

$$P(pass|q^-) = 0.35$$

$$P(pass) = (0.8)(0.7) + (0.35)(0.3) = 0.665$$

So its a simple plug and chug to find:

$$P(q^+)|P(pass) = \frac{(0.8)(0.7)}{0.665} = 0.842$$

$$P(q^-)|P(pass) = \frac{(0.35)(0.3)}{0.665} = 0.158$$

<u>Part c</u> Calculate the optimal decisions given either a pass or a fail, and their expected utilities

If the test is passed, the utility is:

$$(P(q^+)|P(pass))(2000 - 1550) + (P(q^-)|P(pass))(2000 - 2250) = 339.47$$

if it does not pass, then it is:

$$(P(q^+)|P(pass))(2000 - 1550) + (P(q^-)|P(\neg pass))(2000 - 2250) = 42.53$$

<u>Problem 4</u>

For the 43 world shown in Figure 17.1, calculate which squares can be reached from (1,1) by the action sequence[Up,Up,Right] and with what probabili-

ties.Explain how this computation is related to the prediction task (see Section 15.2.1) for a hidden Markov model.

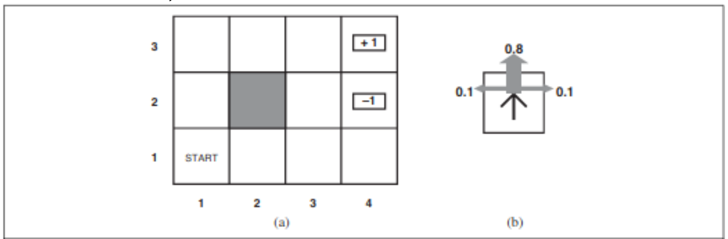

Figure 17.1 (a) A simple 4 × 3 environment that presents the agent with a sequential decision problem. (b) Illustration of the transition model of the environment: the "intended" outcome occurs with probability 0.8, but with probability 0.2 the agent moves at right angles to the intended direction. A collision with a wall results in no movement. The two terminal states have reward +1 and –1, respectively, and all other states have a reward of –0.04.

Answer words

Problem 5

Consider an undiscounted MDP having three states, (1, 2, 3), with rewards1,2,0, respectively. State 3 is a terminal state. In states 1 and 2 there are two possible actions:aandb. The transition model is as follows:
*** In state 1, actionamoves the agent to state 2 with probability 0.8 and makes the agentstay put with probability 0.2.**
*** In state 2, actionamoves the agent to state 1 with probability 0.8 and makes the agentstay put with probability 0.2.**
*** In either state 1 or state 2, actionbmoves the agent to state 3 with probability 0.1 andmakes the agent stay put with probability 0.9.**
Apply policy iteration, showing each step in full, to determine the optimal policy andthe values of states 1 and 2. Assume that the initial policy has action b in both states.

Answer words

Problem 6

Implement the Value-Iteration Algorithm (Figure 17.4 of R&N) and use it to find the optimal utility values and policy for the 4 x 3 world shown in Figure 17.1 of R&N in the case where $R = 1$(and $gamma = 1$).

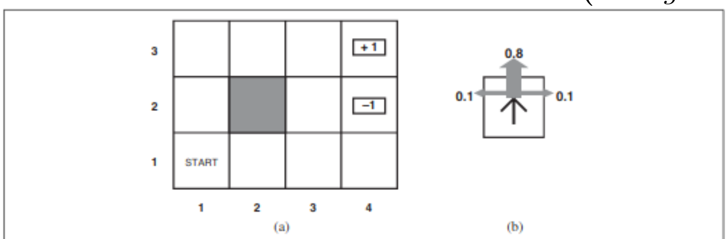

Figure 17.1 (a) A simple 4 × 3 environment that presents the agent with a sequential decision problem. (b) Illustration of the transition model of the environment: the "intended" outcome occurs with probability 0.8, but with probability 0.2 the agent moves at right angles to the intended direction. A collision with a wall results in no movement. The two terminal states have reward +1 and –1, respectively, and all other states have a reward of –0.04.

function VALUE-ITERATION(mdp, ϵ) **returns** a utility function
  inputs: mdp, an MDP with states S, actions $A(s)$, transition model $P(s' \mid s, a)$,
    rewards $R(s)$, discount γ
    ϵ, the maximum error allowed in the utility of any state
  local variables: U, U', vectors of utilities for states in S, initially zero
    δ, the maximum change in the utility of any state in an iteration

  repeat
    $U \leftarrow U'; \delta \leftarrow 0$
    for each state s **in** S **do**
      $U'[s] \leftarrow R(s) + \gamma \max_{a \in A(s)} \sum_{s'} P(s' \mid s, a)\, U[s']$
      if $|U'[s] - U[s]| > \delta$ **then** $\delta \leftarrow |U'[s] - U[s]|$
  until $\delta < \epsilon(1-\gamma)/\gamma$
  return U

Figure 17.4 The value iteration algorithm for calculating utilities of states. The termination condition is from Equation (17.8).

Answer words