Exercise 7.3

(a) Given that he is in state 1, the manufacturer has two possible controls:

$$\mu(1) \in U(1) = \{A : \text{advertise}, \bar{A} : \text{don't advertise}\}$$

Given that he is in state 2, the manufacturer may apply the controls:

$$\mu(2) \in U(2) = \{R : \text{research}, \bar{R} : \text{don't research}\}$$

We want to find an optimal stationary policy, μ, such that Bellman's equation is satisfied. That is, μ should solve:

$$J(i) = \max_{\mu} \mathop{E}_{j}\{g(\mu(i)) + \alpha J(j)\} \qquad i = 1, 2$$

where j is the state following the application of $\mu(i)$ at state i. We can obtain the minimum by solving Bellman's equation for each possible stationary policy and comparing the resulting costs.

For $\mu^1 = (A, R)$:

$$J^1(1) = 4 + \alpha[.8J^1(1) + .2J^1(2)]$$
$$J^1(2) = -5 + \alpha[.7J^1(1) + .3J^1(2)]$$

Letting $\bar{J}^1 = [J^1(1) \quad J^1(2)]'$, we can write:

$$\bar{J}^1 = \begin{bmatrix} 4 \\ -5 \end{bmatrix} + \alpha \begin{bmatrix} .8 & .2 \\ .7 & .3 \end{bmatrix} \bar{J}^1$$

Finally, then:

$$\bar{J}^1 = \left(I - \alpha \begin{bmatrix} .8 & .2 \\ .7 & .3 \end{bmatrix}\right)^{-1} \begin{bmatrix} 4 \\ -5 \end{bmatrix}$$

For $\mu^2 = (A, \bar{R})$, we similarly obtain:

$$\bar{J}^2 = \left(I - \alpha \begin{bmatrix} .8 & .2 \\ .4 & .6 \end{bmatrix}\right)^{-1} \begin{bmatrix} 4 \\ -3 \end{bmatrix}$$

For $\mu^3 = (\bar{A}, R)$:

$$\bar{J}^3 = \left(I - \alpha \begin{bmatrix} .5 & .5 \\ .7 & .3 \end{bmatrix}\right)^{-1} \begin{bmatrix} 6 \\ -5 \end{bmatrix}$$

For $\mu^4 = (\bar{A}, \bar{R})$:

$$\bar{J}^4 = \left(I - \alpha \begin{bmatrix} .5 & .5 \\ .4 & .6 \end{bmatrix}\right)^{-1} \begin{bmatrix} 6 \\ -3 \end{bmatrix}$$

As $\alpha \to 1$, we have for any matrix $M = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$:

$$(I - \alpha M)^{-1} = \frac{1}{(1-\alpha)(1-\alpha+\alpha(p+q))} \begin{bmatrix} 1-\alpha+\alpha q & \alpha p \\ \alpha q & 1-\alpha+\alpha p \end{bmatrix} \to \frac{1}{\delta(p+q)} \begin{bmatrix} q & p \\ q & p \end{bmatrix}$$

where $\delta = 1 - \alpha$. Thus, as $\alpha \to 1$:

$$\bar{J}^1 = \begin{bmatrix} 4 \\ -5 \end{bmatrix}, \quad \bar{J}^2 = \begin{bmatrix} 4 \\ -3 \end{bmatrix}, \quad \bar{J}^3 = \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \quad \bar{J}^4 = \begin{bmatrix} 6 \\ -3 \end{bmatrix}$$

Thus, the optimal stationary policy is the shortsighted one of not advertising or researching.

As $\alpha \to 1$, we have for any matrix $\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$:

$$(I - \alpha M)^{-1} \to \frac{1}{\delta(p+q)} \begin{bmatrix} q & p \\ q & p \end{bmatrix}$$

where $\delta = 1 - \alpha$. Thus, as $\alpha \to 1$:

$$\bar{J}^1 = \frac{1}{\delta} \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \bar{J}^2 = \frac{1}{\delta} \begin{bmatrix} 5/3 \\ 5/3 \end{bmatrix}, \quad \bar{J}^3 = \frac{1}{\delta} \begin{bmatrix} 17/12 \\ 17/12 \end{bmatrix}, \quad \bar{J}^4 = \frac{1}{\delta} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus, the optimal policy is the farsighted one to advertise and research.

(b) Using policy iteration: Let the initial stationary policy be $\mu^0(1) = \bar{A}$ (don't advertise), $\mu^0(1) = \bar{R}$ (don't research). Evaluating this policy yields

$$J_{\mu^0} = (I - \alpha P_{\mu^0})^{-1} g_{\mu^0} = \left(I - 0.9 \begin{bmatrix} .5 & .5 \\ .4 & .6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 6 \\ -3 \end{bmatrix} \approx \begin{bmatrix} 15.49 \\ 5.60 \end{bmatrix}.$$

The new stationary policy satisfying $T_{\mu^1} J_{\mu^0} = T J_{\mu^0}$ is found by solving

$$\mu^1(i) = \arg\max[g(i,u) + \alpha \sum_{j=1}^{2} p_{ij}(u) J_{\mu^0}(j)].$$

We then have

$$\begin{aligned} \mu^1(1) &= \arg\max[4 + 0.9(0.8 J_{\mu^0}(1) + 0.2 J_{\mu^0}(2)), 6 + 0.9(0.5 J_{\mu^0}(1) + 0.5 J_{\mu^0}(2))] \\ &= \arg\max[16.2, 15.5] \\ &= A. \end{aligned}$$

Similarly,

$$\begin{aligned} \mu^1(2) &= \arg\max[-5 + 0.9(0.7 J_{\mu^0}(1) + 0.3 J_{\mu^0}(2)), -3 + 0.9(0.4 J_{\mu^0}(1) + 0.6 J_{\mu^0}(2))] \\ &= \arg\max[6.27, 5.60] \\ &= R. \end{aligned}$$

Evaluating this new policy yields

$$J_{\mu^1} = \left(I - 0.9 \begin{bmatrix} .8 & .2 \\ .7 & .3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 4 \\ -5 \end{bmatrix} \approx \begin{bmatrix} 22.20 \\ 12.31 \end{bmatrix}.$$

Attempting to find another improved policy, we see that

$$\begin{aligned} \mu^2(1) &= \arg\max[4 + 0.9(0.8 J_{\mu^1}(1) + 0.2 J_{\mu^1}(2)), 6 + 0.9(0.5 J_{\mu^1}(1) + 0.5 J_{\mu^1}(2))] \\ &= \arg\max[22.20, 21.53] \\ &= A, \end{aligned}$$

and

$$\begin{aligned}\mu^2(2) &= \arg\max[-5 + 0.9(0.7J_{\mu^1}(1) + 0.3J_{\mu^1}(2)), -3 + 0.9(0.4J_{\mu^1}(1) + 0.6J_{\mu^1}(2))] \\ &= \arg\max[12.31, 11.64] \\ &= R.\end{aligned}$$

Since $J_{\mu^1} = TJ_{\mu^1}$, we're done. The optimal policy is thus $\mu = (A, R)$.

The linear programming formulation for this problem is

$$\min \lambda_1 + \lambda_2$$

subject to

$$\begin{aligned}\lambda_1 &\geq 4 + 0.9[0.8\lambda_1 + 0.2\lambda_2] \\ \lambda_1 &\geq 6 + 0.9[0.5\lambda_1 + 0.5\lambda_2] \\ \lambda_2 &\geq -5 + 0.9[0.7\lambda_1 + 0.3\lambda_2] \\ \lambda_2 &\geq -3 + 0.9[0.4\lambda_1 + 0.6\lambda_2].\end{aligned}$$

By plotting these equations or by using an LP package, we see that the optimal costs are $J^*(1) = \lambda_1^* = 22.20$ and $J^*(2) = \lambda_2^* = 12.31$.

Exercise 7.5

(a) Define three states: $\{(s,r)$: the umbrella is in the same location as the person and it is raining, (s,n) : the umbrella is in the same location as the person and it is not raining, and o : the umbrella is in the other location$\}$. In state (s,n), the person makes the decision whether or not to take the umbrella. In state (s,r), the person has no choice and takes the umbrella. In state o, the person also has no choice and does not take the umbrella. Bellman's equation yields

$$J(o) = pW + \alpha pJ(s,r) + \alpha(1-p)J(s,n)$$

$$J(s,r) = \alpha pJ(s,r) + \alpha(1-p)J(s,n)$$

$$J(s,n) = \min[\alpha J(o), V + \alpha pJ(s,r) + \alpha(1-p)J(s,n).$$

An alternative is to use the following two states are: $\{s$: the umbrella is in the same location as the person, o : the umbrella is in the other location$\}$. In state s, the person takes the umbrella with probability p (if it rains) and makes a decision whether or not to take the umbrella with probability $1-p$ (if it doesn't rain). In state o, the person has no decision to make. Bellman's equation yields

$$J(o) = pW + \alpha J(s)$$

$$\begin{aligned} J(s) &= p\alpha J(s) + (1-p)\min[V + \alpha J(s), \alpha J(o)] \\ &= \min[(1-p)V + \alpha J(s), p\alpha J(s) + (1-p)\alpha J(o)]. \end{aligned}$$

(b) In the two-state formulation, since $J(o)$ is a linear function of $J(s)$, we need only concentrate on minimizing $J(s)$. The two possible stationary policies are $\mu^1(s) = \{T$: take umbrella$\}$ and $\mu^2(s) = \{L$: leave umbrella$\}$.

For μ^1, we have

$$\begin{aligned} J^1(s) &= (1-p)V + \alpha J(s) \\ &= \frac{(1-p)V}{1-\alpha}. \end{aligned}$$

For μ^2, we have

$$\begin{aligned} J^2(s) &= p\alpha J(s) + (1-p)\alpha J(o) \\ &= p\alpha J(s) + (1-p)\alpha[pW + \alpha J(s)] \\ &= \frac{(1-p)pW}{\frac{1}{\alpha} - p - (1-p)\alpha}. \end{aligned}$$

So the optimal policy is to take the umbrella whenever possible if

$$J^1(s) < J^2(s),$$

or when

$$\frac{(1-p)V}{1-\alpha} < \frac{(1-p)pW}{\frac{1}{\alpha} - p - (1-p)\alpha}.$$

This expression simplifies to

$$p > \frac{\frac{V}{\alpha}(1+\alpha)}{W+V}.$$

Using the three-state formulation, we see from the second equation that

$$J(s,n) = \frac{1-\alpha p}{\alpha(1-p)} J(s,r).$$

Then, the other two equations become

$$J(o) = pW + J(s,r)$$

and

$$J(s,n) = \min[\alpha J(o), V + J(s,r)].$$

$J(o)$ and $J(s,n)$ are linear functions of $J(s,r)$ so again, we can just concentrate on minimizing $J(s,r)$ via the equation

$$\frac{1-\alpha p}{\alpha(1-p)} J(s,r) = \min[\alpha J(o), V + J(s,r)].$$

Using the same process as in the two-state formulation, we get the same result.

Exercise 7.7

Suppose that $J_k(i+1) \geq J_k(i)$ for all i. We will show that $J_{k+1}(i+1) \geq J_{k+1}(i)$ for all i. Consider first the case $i+1 < n$. Then by the induction hypothesis, we have

$$c(i+1) + \alpha(1-p)J_k(i+1) + \alpha p J_k(i+2) \geq ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1). \tag{1}$$

Define for any scalar γ,

$$F_k(\gamma) = \min\big[K + \alpha(1-p)J_k(0) + \alpha p J_k(1),\ \gamma\big].$$

Since $F_k(\gamma)$ is monotonically increasing in γ, we have from Eq. (1),

$$\begin{aligned} J_{k+1}(i+1) &= F_k\big(c(i+1) + \alpha(1-p)J_k(i+1) + \alpha p J_k(i+2)\big) \\ &\geq F_k\big(ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1)\big) \\ &= J_{k+1}(i). \end{aligned}$$

Finally, consider the case $i+1 = n$. Then, we have

$$\begin{aligned} J_{k+1}(n) &= K + \alpha(1-p)J_k(0) + \alpha p J_k(1) \\ &\geq F_k\big(ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1)\big) \\ &= J_{k+1}(n-1). \end{aligned}$$

The induction is complete.

Exercise 7.8

A threshold policy is specified by a threshold integer m and has the form

Process the orders if and only if their number exceeds m.

The cost function corresponding to a threshold policy specified by m will be denoted by J_m. By Prop. 3.1(c), this cost function is the unique solution of system of equations

$$J_m(i) = \begin{cases} K + \alpha(1-p)J_m(0) + \alpha p J_m(1) & \text{if } i > m, \\ ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) & \text{if } i \leq m. \end{cases} \tag{1}$$

Thus for all $i \leq m$, we have

$$J_m(i) = \frac{ci + \alpha p J_m(i+1)}{1 - \alpha(1-p)},$$

$$J_m(i-1) = \frac{c(i-1) + \alpha p J_m(i)}{1 - \alpha(1-p)}.$$

From these two equations it follows that for all $i \leq m$, we have

$$J_m(i) \leq J_m(i+1) \quad \Rightarrow \quad J_m(i-1) < J_m(i). \tag{2}$$

Denote now

$$\gamma = K + \alpha(1-p)J_m(0) + \alpha p J_m(1).$$

Consider the policy iteration algorithm, and a policy $\overline{\mu}$ that is the successor policy to the threshold policy corresponding to m. This policy has the form

Process the orders if and only if

$$K + \alpha(1-p)J_m(0) + \alpha p J_m(1) \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1)$$

or equivalently

$$\gamma \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1).$$

In order for this policy to be a threshold policy, we must have for all i

$$\gamma \leq c(i-1) + \alpha(1-p)J_m(i-1) + \alpha p J_m(i) \quad \Rightarrow \quad \gamma \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1). \tag{3}$$

This relation holds if the function J_m is monotonically nondecreasing, which from Eqs. (1) and (2) will be true if $J_m(m) \leq J_m(m+1) = \gamma$.

Let us assume that the opposite case holds, where $\gamma < J_m(m)$. For $i > m$, we have $J_m(i) = \gamma$, so that

$$ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) = ci + \alpha\gamma. \tag{4}$$

We also have

$$J_m(m) = \frac{cm + \alpha p \gamma}{1 - \alpha(1-p)},$$

from which, together with the hypothesis $J_m(m) > \gamma$, we obtain

$$cm + \alpha\gamma > \gamma. \tag{5}$$

Thus, from Eqs. (4) and (5) we have

$$ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) > \gamma, \qquad \text{for all } i > m, \tag{6}$$

so that Eq. (3) is satisfied for all $i > m$.

For $i \le m$, we have $ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) = J_m(i)$, so that the desired relation (3) takes the form

$$\gamma \le J_m(i-1) \quad \Rightarrow \quad \gamma \le J_m(i). \tag{7}$$

To show that this relation holds for all $i \le m$, we argue by contradiction. Suppose that for some $i \le m$ we have $J_m(i) < \gamma \le J_m(i-1)$. Then since $J_m(m) > \gamma$, there must exist some $\bar{i} > i$ such that $J_m(\bar{i}-1) < J_m(\bar{i})$. But then Eq. (2) would imply that $J_m(j-1) < J_m(j)$ for all $j \le \bar{i}$, contradicting the relation $J_m(i) < \gamma \le J_m(i-1)$ assumed earlier. Thus, Eq. (7) holds for all $i \le m$ so that Eq. (3) holds for all i. The proof is complete.

Exercise 7.10

(a) The states are s^i, $i = 1, \ldots, n$, corresponding to the worker being unemployed and being offered a salary w^i, and $\bar{s}^i$, $i = 1, \ldots, n$, corresponding to the worker being employed at a salary level w^i. Bellman's equation is

$$J(s^i) = \max\left[c + \alpha \sum_{j=1}^{n} \xi_j J(s^j),\ w^i + \alpha J(\bar{s}^i)\right], \qquad i = 1, \ldots, n, \tag{1}$$

$$J(\bar{s}^i) = w^i + \alpha J(\bar{s}^i), \qquad i = 1, \ldots, n, \tag{2}$$

where ξ_j is the probability of an offer at salary level w^j at any one period.

From Eq. (2), we have

$$J(\bar{s}^i) = \frac{w^i}{1-\alpha} \qquad i = 1, \ldots, n,$$

so that from Eq. (1) we obtain

$$J(s^i) = \max\left[c + \alpha \sum_{j=1}^{n} \xi_j J(s^j),\ \frac{w^i}{1-\alpha}\right],$$

Thus it is optimal to accept salary w^i if

$$w^i \ge (1-\alpha)\left(c + \alpha \sum_{j=1}^{n} \xi_j J(s^j)\right).$$

The right-hand side of the above relation gives the threshold for acceptance of an offer.

(b) In this case Bellman's equation becomes

$$J(s^i) = \max\left[c + \alpha \sum_{j=1}^{n} \xi_j J(s^j),\ w^i + \alpha\left((1-p_i)J(\bar{s}^i) + p_i \sum_{j=1}^{n} \xi_j J(s^j)\right)\right] \tag{3}$$

$$J(\bar{s}^i) = w^i + \alpha\left((1-p_i)J(\bar{s}^i) + p_i \sum_{j=1}^{n} \xi_j J(s^j)\right). \tag{4}$$

Let us assume without loss of generality that

$$w^1 < w^2 < \cdots < w^n.$$

Let us assume further that $p_i = p$ for all i. From Eq. (4), we have

$$J(\bar{s}^i) = \frac{w^i + p\sum_{j=1}^n \xi_j J(s^j)}{1 - \alpha(1-p)},$$

so it follows that

$$J(\bar{s}^1) < J(\bar{s}^2) < \cdots < J(\bar{s}^n). \tag{5}$$

We thus obtain that the second term in the maximization of Eq. (3) is monotonically increasing in i, implying that there is a salary threshold above which the offer is accepted.

In the case where p_i is not independent of i, salary level is not the only criterion of choice. There must be consideration for job security (the value of p_i). However, if p_i and w^i are such that Eq. (5) holds, then there still is a salary threshold above which the offer is accepted.

Exercise 7.11

Using the notation of Exercise 7.10, Bellman's equation has the form

$$\lambda + h(s^i) = \max\left[c + \sum_{j=1}^n \xi_j h(s^i),\ w^i + (1-p_i)h(\bar{s}^i) + p_i \sum_{j=1}^n \xi_j h(s^j)\right], \qquad i = 1,\ldots,n, \tag{3}$$

$$\lambda + h(\bar{s}^i) = w^i + (1-p_i)h(\bar{s}^i) + p_i \sum_{j=1}^n \xi_j h(s^j), \qquad i = 1,\ldots,n. \tag{4}$$

From these equations, we have

$$\lambda + h(s^i) = \max\left[c + \sum_{j=1}^n \xi_j h(s^j),\ \lambda + h(\bar{s}^i)\right], \qquad i = 1,\ldots,n,$$

so it is optimal to accept a salary offer w^i if $h(\bar{s}^i)$ is no less that the threshold

$$c - \lambda + \sum_{j=1}^n \xi_j h(s^j).$$

Here λ is the optimal average salary per period (over an infinite horizon). If $p_i = p$ for all i and $w^1 < w^2 < \cdots < w^n$, then from Eq. (4) it follows that $h(\bar{s}^i)$ is monotonically increasing in i, and the optimal policy is to accept a salary offer if it exceeds a certain threshold.