Problem 1 (15 points)

(a) Let μ_A (μ_B) be the stationary control applied when in town A (B). The control $\mu \in \{Stay, Change\}$. We can obtain the optimal stationary control by solving Bellman's equation for each of the four possible policies. Let $\mu = (\mu_A, \mu_B)$.

For $\mu^1 = (S, S)$:

$$J_A^1 = r_A + \alpha J_A^1$$
$$J_B^1 = r_B + \alpha J_B^1$$

So,

$$J_A^1 = \frac{r_A}{1-\alpha}; \qquad J_B^1 = \frac{r_B}{1-\alpha}$$

For $\mu^2 = (S, C)$:

$$J_A^2 = \frac{r_A}{1-\alpha}$$
$$J_B^2 = r_a - c + \alpha J_A^2 = -c + \frac{r_A}{1-\alpha}$$

For $\mu^3 = (C, S)$:

$$J_A^3 = -c + \frac{r_B}{1-\alpha}; \qquad J_B^3 = \frac{r_B}{1-\alpha}$$

For $\mu^4 = (C, C)$:

$$J_A^4 = r_A - c + \alpha J_B^4$$
$$J_B^4 = r_B - c + \alpha J_A^4$$

Thus,

$$J_A^4 = \frac{r_A + \alpha r_B - (1+\alpha)c}{1-\alpha^2}; \qquad J_B^4 = \frac{r_B + \alpha r_A - (1+\alpha)c}{1-\alpha^2}$$

As $\alpha \to 0$, $\bar{J}^1 = \begin{bmatrix} J_A^1 \\ J_B^1 \end{bmatrix} = \begin{bmatrix} r_A \\ r_B \end{bmatrix}$ is clearly optimal. Thus, the optimal policy is for the salesmen to stay in the town he starts in. As $\alpha \to 1$, we have:

$$(1-\alpha)\bar{J}^1 = \begin{bmatrix} r_A \\ r_B \end{bmatrix}, \qquad (1-\alpha)\bar{J}^2 = \begin{bmatrix} r_A \\ r_A \end{bmatrix}$$
$$(1-\alpha)\bar{J}^3 = \begin{bmatrix} r_B \\ r_B \end{bmatrix}, \qquad (1-\alpha)\bar{J}^4 = \begin{bmatrix} \frac{r_A}{2} - c \\ \frac{r_B}{2} - c \end{bmatrix}$$

Since $c > r_A > r_B$, μ^2 is optimal. That is, the salesman should move to A and remain there.

(b) For $c = 3$, $r_A = 2$, $r_B = 1$, and $\alpha = .9$:

$$J^1 = \begin{bmatrix} 20 \\ 10 \end{bmatrix}, \quad J^2 = \begin{bmatrix} 20 \\ 17 \end{bmatrix}, \quad J^3 = \begin{bmatrix} 7 \\ 10 \end{bmatrix}, \quad J^4 = \begin{bmatrix} -15.26 \\ -14.74 \end{bmatrix}$$

Thus, the optimal policy is to move into A and remain there.

Problem 2 (50 points)

a) (i) First, we need to define a state space for the problem. The obvious choice for a state variable is our location. However, this does not encapsulate all of the necessary information. We also need to include the value of c if it is known. Thus, let the state space consist of the following $2m+2$ states: $\{S, S_1, \ldots, S_m, I_1, \ldots I_m, D\}$, where S is associated with being at the starting point with no information, S_i and I_i are associated with being at S and I, respectively, and knowing that $c = c_i$, and D is the termination state.

At state S, there are two possible controls: go directly to D (*direct*) or go to an intermediate point (*indirect*). If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$.

At state S_i, for $i \in \{1, \ldots, m\}$, we have the same controls as at state S. Again, if control *direct* is selected, we go to state D with probability 1, and the cost is $g(S_i, direct, D) = a$. If, on the other hand, control *indirect* is selected, we go to state I_i with probability 1, and the cost is $g(S, indirect, I_i) = b$.

At state I_i, for $i \in \{1, \ldots, m\}$, there are also two possible controls: go back to the start (*start*) or go to the destination (*dest*). If control *start* is selected, we go to state S_i with probability 1, and the cost is $g(I_i, start, S_i) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is $g(I_i, dest, D) = c_i$.

We have thus formulated the problem as a stochastic shortest path problem. Bellman's equation for this problem is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(S_i) = \min[a, b + J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S_i)].$$

We assume that $b > 0$. Then, Assumptions 5.1 and 5.2 hold since all improper policies have infinite cost. As a result, if $\mu^*(I_i) = start$, then $\mu^*(S_i) = direct$. If $\mu^*(I_i) \neq start$, then we never reach state S_i and so it doesn't matter what the control is in this case. Thus, $J^*(S_i) = a$, and $\mu^*(S_i) = direct$. From this, it is easy to derive the optimal costs and controls for the other states:

$$J^*(I_i) = \min[c_i, b + a] \qquad \mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < b + a \\ \text{start}, & \text{otherwise,} \end{cases}$$

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i \min(c_i, b + a)]$$

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i \min(c_i, b + a) \\ \text{indirect}, & \text{otherwise.} \end{cases}$$

For the numerical case given, we see that $a < b + \sum_{i=1}^{m} p_i \min(c_i, b+a)$ since $a = 2$ and $b + \sum_{i=1}^{m} p_i \min(c_i, b + a) = 2.5$. Hence $\mu(S) = direct$. We need not consider the other states since they will never be reached.

(ii) In this case, every time we are at the starting location, our available information is the same. We thus no longer need the states S_i from part (i). Our state space for this part is then $S, I_1, \ldots, I_m, D$.

At state S, the possible controls are $\{direct, indirect\}$. If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$ [same as in part (ii)].

At state I_i, for $i \in \{1, \ldots, m\}$, the possible controls are $\{start, dest\}$. If control *start* is selected, we go to state S with probability 1, and the cost is $g(I_i, start, S) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is $g(I_i, dest, D) = c_i$.

Bellman's equation for this stochastic shortest path problem is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S)].$$

The optimal policy can be described by

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i J^*(I_i) \\ \text{indirect}, & \text{otherwise,} \end{cases}$$

$$\mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < b + J^*(S) \\ \text{start}, & \text{otherwise.} \end{cases}$$

We will solve the problem for the numerical case by "guessing" an optimal policy and then showing that the resulting cost J_μ satisfies $J = TJ$. Since J^* is the unique solution to this equation, our policy is optimal. So let's guess the initial policy to be

$$\mu^*(S) = direct \qquad \mu^*(I_1) = dest \qquad \mu^*(I_2) = start.$$

Then

$$J(S) = a = 2 \qquad J(I_1) = c_1 = 0 \qquad J(I_2) = b + J^*(S) = 1 + 2 = 3.$$

From Bellman's equation, we have

$$J(S) = \min(2, 1 + 0.5(3 + 0)) = 2$$

$$J(I_1) = \min(0, 1 + 2)) = 0$$

$$J(I_2) = \min(5, 1 + 2)) = 3.$$

Thus, our policy is optimal.

b) The state space for this problem is the same as for part a(ii): $\{S, I_1, \ldots, I_m, D\}$.

At state S, the possible controls are $\{direct, indirect\}$. If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$ [same as in part a,(i) and (ii)].

At state I_i, for $i \in \{1, \ldots, m\}$, we have an additional option of waiting. So the possible controls are $\{start, dest, wait\}$. If control *start* is selected, we go to state S with probability 1, and the cost is $g(I_i, start, S) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is

$g(I_i, dest, D) = c_i$. If control *wait* is selected, we go to state I_j with probability p_j, and the cost is $g(I_i, wait, I_j) = d$.

Bellman's equation is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S), d + \sum_{j=1}^{m} p_j J^*(I_j)].$$

We can describe the optimal policy as follows:

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i J^*(I_i) \\ \text{indirect}, & \text{otherwise.} \end{cases}$$

If *direct* was selected, we do not need to consider the other states (other than D) since they will never be reached. If *indirect* was selected, then defining $k = \min(2b, d)$, we see that

$$\mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < k + \sum_{i=1}^{m} J^*(I_i) \\ \text{start}, & \text{if } c_i > k + \sum_{i=1}^{m} J^*(I_i) \text{ and } 2b < d \\ \text{wait}, & \text{if } c_i > k + \sum_{i=1}^{m} J^*(I_i) \text{ and } 2b > d. \end{cases}$$

Problem 3 (35 points)

Let the state be the current set of wins.

(a)

$$J^*(0) = \min_i \{c_i + p_i J^*(i) + (1 - p_i) J^*(0)\}$$

$$J^*(i) = \min_{j \neq i} \{c_j + p_j J^*(i, j) + (1 - p_j) J^*(0)\}$$

$$J^*(i, j) = c_k - p_k m + (1 - p_k) J^*(0) \qquad k \neq i, j$$

(b) Let J represent the cost-to-go for the stationary policy ijk. Then J satisfies the following set of equations:

$$J(0) = c_i + p_i J(i) + (1 - p_i) J(0)$$

$$J(i) = c_j + p_j J(i, j) + (1 - p_j) J(0)$$

$$J(i, j) = c_k - p_k m + (1 - p_k) J(0)$$

Solving the above equations for $J(0)$, the expected cost of policy ijk, we have:

$$J(0) = \frac{c_i + p_i c_j + p_i p_j c_k - p_i p_j p_k m}{p_i p_j p_k}$$

(c)

$$\lambda^* + h^*(0) = \min_i \{c_i + p_i h^*(i) + (1 - p_i) h^*(0)\}$$

$$\lambda^* + h^*(i) = \min_{j \neq i} \{c_j + p_j h^*(i, j) + (1 - p_j) h^*(0)\}$$

$$\lambda^* + h^*(i, j) = c_k + p_k(-m + h^*(0)) + (1 - p_k) h^*(0) \qquad k \neq i, j$$

(d) Let λ represent the average cost-per-stage for the stationary policy ijk. Then λ satisfies the following set of equations:

$$\lambda + h(0) = c_i + p_i h(i) + (1 - p_i) h(0)$$

$$\lambda + h(i) = c_j + p_j h(i, j) + (1 - p_j) h(0)$$

$$\lambda + h(i, j) = c_k + p_k(-m + h(0)) + (1 - p_k) h(0) \qquad k \neq i, j$$