Math 312 Portfolio

Delaney Scheiern

Part I

Problem

A baseball team's bullpen, or the group of pitchers, is one of the most important parts of a Major League Baseball (MLB) team. Without strong pitchers, a team will never be able to win, even if they have the best batters. My hometown MLB team, the Angels, has struggled with pitching in recent years. I'm interested in trying to improve their pitching, in order to optimize their likelihood of winning games. I will try to do this by adjusting the innings in which four predetermined pitchers will pitch. I was not able to find a model that has tried to solve this problem, but team managers are constantly considering their options by looking at past data and experiences to choose the pitching order.

Approach

I will be using a Markov Decision Chain with two absorbing states — win (W) and lose (L). I will try to optimize the long run probability of ending at a win rather than a loss. Each of the states will consist of three parts — pitcher number, inning number, and whether the team is winning (W), losing (L), or tied (T). I will be focusing on earned run average (ERA), which is the mean of earned runs given up by a pitcher per nine innings pitched, and walks plus hits per inning pitched (WHIP), which measures the number of baserunners a pitcher has allowed per inning pitched. I will also account for a pitcher's pitching limit in the assumptions.

Assumptions

I am going to assume:

- There are always only 9 innings in a game
- Pitchers cannot be changed mid-inning
- All teams will have 4 pitchers in a designated order for each game
    - Pitchers can only go in order (1,2,3,4) but can be skipped
    - 1st pitcher is a starter (up to ~100 pitches per game)
    - 2nd and 3rd pitchers are long relievers (up to ~50 pitches per game)
    - 4th pitcher is a reliever (up to ~20 pitches per game)

- All batters perform the same, at the average of all batters in MLB
- Pitchers will always perform at their personal average, no matter the day or opponent
- Once a pitcher has pitched and is taken out of the game, they cannot pitch again that game

I also will make assumptions about transitions, which are represented in mini transition matrices:

Transitioning to L/T/W from L/T/W

	L	**T**	**W**
L	0.5	-0.5	-1.5
T	-0.5	0.5	-0.5
W	-1.5	-0.5	0.5

These values represent the weight of transitioning from losing, tied, or winning states to losing, tied, or winning states. They will be multiplied by the positive scalar a, which is one of the values to be chosen during optimization. I am assuming that a game is more likely to stay in its current L/T/W state than to transition to another state. I am also assuming that it is more difficult to transition from losing to winning or winning to losing than going to and from being tied.

Transitioning Pitchers Based on W/L

	1	**2**	**3**	**4**
1	If W, +1; if L, -1	X	X	X
2	If W, -1; if L, +3	If W, +1; if L, -1	X	X
3	If W, -2; if L, +2	If W, -1; if L, +2	If W, +1; if L, -1	X
4	If W, -3; if L, +1	If W, -2; if L, +1	If W, -1; if L, +1	1

These positive and negative values are the weights that will be multiplied by the positive scalar b, which is one of the values to be chosen during optimization. In this matrix, I assume that if the team is winning they are less likely to change pitchers, while if they are losing they are more likely to change pitchers. I also assumed that it is less likely for a team to skip over a pitcher than to go to the next one in the lineup.

Transitioning Pitchers Based on Inning

	1	**2**	**3**	**4**
1	If 1-5: +1; if 6-8: -1	X	X	X
2	If 1-4,7,8: -1; if 5-7: +1	If 1-4,7,8: -1; if 5-7: +1	X	X
3	If 1-5,8: -1; if 6,7: +1	If 1-5,8: -1; if 6,7: +1	If 1-5,8: -1; if 6,7: +1	X
4	If 1-7: -1; if 8: +1	If 1-7: -1; if 8: +1	If 1-7: -1; if 8: +1	1

These positive and negative values will be multiplied by the positive scalar c, which is one of the values to be chosen during optimization. This matrix attempts to account for a pitcher's maximum number of pitches per game. I assumed that the starting pitcher is likely to stay in at least the first six innings, since they often can pitch up to 100 pitches. After these 100 pitches, typically around the seventh inning, it is more likely for a new pitcher to come into the game. If the game is going into the ninth inning, it is likely that the fourth pitcher, the best reliever, will join the game even if the second and third pitchers never played.

Transition Matrix

To determine the transition matrix for this Markov Decision Process, I will account for all of the above assumptions, as well as including ERA and WHIP. To incorporate ERA and WHIP, I will measure the distance of each pitcher's ERA and WHIP from the MLB average values. Let $aERA = average\ ERA\ in\ MLB$, $ERA = individual\ pitcher's\ ERA$, $aWHIP = average\ WHIP\ in\ MLB$, and $WHIP = individual\ pitcher's\ WHIP$. Since a lower ERA and WHIP is considered better, $aERA - ERA > 0$ if the pitcher has a better than average ERA, and $aWHIP - WHIP > 0$ if the pitcher has a better than average WHIP. Each of these values will be either added or subtracted from each transition probability after being multiplied by a positive scalar — d for ERA and e for WHIP.

A slightly more complete transition matrix can be referenced in the Baseball Markov Chain Matrix PDF, however I will demonstrate and explain the probability of transitioning from state $1, 2, W$ to all of its possible states.

	1,2,W
1,3,W	$p1 = 0.5a + b + c + d(aERA - ERA) + e(aWHIP - WHIP)$
1,3,L	$p2 = -1.5a + b + c - d(aERA - ERA) - e(aWHIP - WHIP)$
1,3,T	$p3 = -0.5a + b + c - 0.5d(aERA - ERA) - 0.5e(aWHIP - WHIP)$
2,3,W	$p4 = 0.5a - b - c + d(aERA - ERA) + e(aWHIP - WHIP)$
2,3,L	$p5 = -1.5a - b - c - d(aERA - ERA) - e(aWHIP - WHIP)$
2,3,T	$p6 = -0.5a - b - c - 0.5d(aERA - ERA) - 0.5e(aWHIP - WHIP)$
3,3,W	$p7 = 0.5a - 2b - c + d(aERA - ERA) + e(aWHIP - WHIP)$
3,3,L	$p8 = -1.5a - 2b - c - d(aERA - ERA) - e(aWHIP - WHIP)$
3,3,T	$p9 = -0.5a - 2b - c - 0.5d(aERA - ERA) - 0.5e(aWHIP - WHIP)$
4,3,W	$p10 = 0.5a - 3b - c + d(aERA - ERA) + e(aWHIP - WHIP)$
4,3,L	$p11 = -1.5a - 3b - c - d(aERA - ERA) - e(aWHIP - WHIP)$
4,3,T	$p12 = -0.5a - 3b - c - 0.5d(aERA - ERA) - 0.5e(aWHIP - WHIP)$

These equations take some inspiration from dynamical systems, such as predator/prey models. If a value is subtracted, it is harming the “population,” or probability in this case. If a value is added, it is helping the probability. With respect to ERA and WHIP, if the pitcher is better than average, the probability of winning should be increased, while the probability of dropping to tied or losing should be decreased.

Optimization

Once all of the equations are completed to fill the transition matrix, we can begin trying to optimize the probability of being absorbed at W. Since this is an absorbing chain, we can write the transition matrix in the form

$$\left[\frac{I \mid R}{0 \mid Q}\right]$$

And

$$\lim_{k\to\infty} T^k = \left[\frac{I \mid RN}{0 \mid 0}\right] \text{ where } N = (I - Q)^{-1}$$

Let $B = RN$ where B_{mj} is the probability of being absorbed at absorbing state m if we start at state j. B_{11} will be the probability of being absorbed at win. This cell will have the equation we want to maximize in terms of a, b, c, d, and e. We also have our constraint equations, because each column in the transition matrix sums to 1. We can use the method of Lagrange multipliers to maximize the equation. Let $B_{11} = f(a,b,c,d,e)$. Let $g_k(a,b,c,d,e)$ be the sums of each of the columns. Then solving $grad(f) = \sum_{k=1}^{m} \lambda_k \cdot grad(g_k)$ for each Lagrange multiplier, we get the differential equations:

$$\frac{\partial f}{\partial a} = \lambda_1 \frac{\partial g_1}{\partial a} + \lambda_2 \frac{\partial g_2}{\partial a} + \cdots + \lambda_m \frac{\partial g_m}{\partial a}$$
$$\frac{\partial f}{\partial b} = \lambda_1 \frac{\partial g_1}{\partial b} + \lambda_2 \frac{\partial g_2}{\partial b} + \cdots + \lambda_m \frac{\partial g_m}{\partial b}$$
$$\frac{\partial f}{\partial c} = \lambda_1 \frac{\partial g_1}{\partial c} + \lambda_2 \frac{\partial g_2}{\partial c} + \cdots + \lambda_m \frac{\partial g_m}{\partial c}$$
$$\frac{\partial f}{\partial d} = \lambda_1 \frac{\partial g_1}{\partial d} + \lambda_2 \frac{\partial g_2}{\partial d} + \cdots + \lambda_m \frac{\partial g_m}{\partial d}$$
$$\frac{\partial f}{\partial e} = \lambda_1 \frac{\partial g_1}{\partial e} + \lambda_2 \frac{\partial g_2}{\partial e} + \cdots + \lambda_m \frac{\partial g_m}{\partial e}$$

Using these equations, we can obtain the optimal values of the variables to maximize the team’s chance of winning.

Possible Expansions

There are a lot of opportunities to improve or expand on this model. Since the transition matrix is too large to calculate by hand, it would be best to create a computer program to do the calculations for us. Particularly, a neural network system would work very well with this model, as it could be given the expansive baseball data as training data and be able to tweak the variables until it found the optimal choices.

Another way this model could be improved is to take more variables into consideration. I chose ERA and WHIP as my two statistics because they are generally agreed upon as most important, but there are tons of other options for evaluating pitchers. If we used as many of these statistics as possible and used the neural network system described above, the computer program could tell us which statistics are actually most important.

As always, by adjusting the assumptions I made, the model can be drastically changed. For example, teams do not typically have a stagnant pitching lineup for each game, and it is fairly common for baseball games to go over nine innings. These tweaks could improve the accuracy of the model.

One of the most interesting aspects of this topic is the vast amount of baseball data that has been collected over the last century. By using this data, either in the more simple model I created or as training data for the neural network program I proposed, we could analyze pitching strategies from the beginning of MLB and compare them to today’s game.

Part II

Theorem

$T^{(k)} = T^k$ i.e. to get the k-step transition matrix of a Markov chain, simply raise the original matrix to the k^{th} power.

A Markov chain describes the probability of transitioning between different states, solely based on the current state. The transition matrix of a Markov chain is column stochastic, so all the entries are positive and each column sums to 1.

This theorem is relevant to many of the problems we encountered with Markov chains. For example, we could compute the probabilities of recidivism and what the distribution was after a certain timespan. It is also important for Part I of my portfolio, in that we could not optimize the outcomes of a Markov chain without knowing what the probabilities will be at future steps. It is also the basis for finding the long run probabilities of a Markov chain, which I believe is one of the most powerful aspects of Markov chains. Just by knowing the beginning states, we can predict the future states using this theorem.

Proof

This is a proof by induction.

Let $k = 1$. Then by definition the transition matrix of a Markov chain to the first power is equal to the 1-step transition matrix.

Let $k > 1$.

$$T_{ij}^{(k)} = Prob(R_{n+k} = S_j \mid R_n = S_i)$$

$$= \frac{Prob(R_{n+k}=S_j,\, R_n=S_i)}{Prob(R_n=S_i)}$$

$$= \sum_{l=1}^{N} \frac{Prob(R_{n+k}=S_j,\, R_{n+k-1}=S_l,\, R_n=S_i)}{Prob(R_n=S_i)}$$

$$= \sum_{l=1}^{N} \frac{Prob(R_{n+k}=S_j,\, R_{n+k-1}=S_l,\, R_n=S_i)}{Prob(R_{n+k-1}=S_l,\, R_n=S_i)} \cdot \frac{Prob(R_{n+k-1}=S_l,\, R_n=S_i)}{Prob(R_n=S_i)}$$

$$= \sum_{l=1}^{N} Prob(R_{n+k} = S_j \mid R_{n+k-1} = S_k,\ R_n = S_i) \cdot (T^{k-1})_{il}$$

$$= \sum_{l=1}^{N} T_{lj} \cdot (T^{k-1})_{il}$$

$$= (T^k)_{ij}$$

Thus for any transition matrix, the probabilities of the k-step transition matrix are equal to the original transition matrix raised to the k^{th} power.