ANALYSIS OF A BASEBALL SIMULATION GAME USING MARKOV CHAINS

DONALD M. DAVIS

1. Introduction

APBA baseball is a baseball simulation game invented by Dick Seitz of Lancaster, Pennsylvania, and first marketed in 1951. It has an avid following today([1]), and was featured in a *New York Times* article in August 2009.([5]) In this article, we use Markov chains to analyze certain aspects of this game. For example, we can tell whether one player's batting card is better than another's, and we can make informed decisions about strategy in the game.

Markov chain analysis has been applied to real baseball in [2], [6], and [7]. For example, two of these discuss how to optimize the batting order, using Markov chains and detailed statistics available from a database. The rigid structure of the game analyzed here lends itself to conclusions of a different nature than those.

Each play of the game begins with a roll of a pair of dice, one red and one white. A red 4 and white 2 is interpreted as 42. There are $6 \cdot 6 = 36$ possible rolls. Each player has a card, which is based on his performance during a particular season. The card associates to each of the 36 possible dice rolls a number from 1 to 41. You look on the batter's card to see what number corresponds to the number rolled. For example, on most players' cards the number corresponding to a roll of 42 will either be 13 or 14. Then one looks in a book to see what is the outcome of this number. This will depend on which bases are occupied, perhaps on the opposing pitcher and fielders, perhaps on the speed of the baserunners, and occasionally on how many are out. Usually 13 is a strikeout and 14 is a walk.

The value of a player's card depends on the 36 numbers on it, since one may assume that each of the 36 numbers is rolled equally often. We can compute mathematically the expected number of runs that would be scored from any (base,out) situation by a team of average players, and then, using this information, we can determine the average increase or decrease in expected number of runs scored when any number comes up on a player's card. For example, 1 is always a home run. We determine that when a 1 comes up, the team's expected number of runs for that inning is increased by 1.41. On the other hand, 13, which is usually a strikeout, decreases the team's expected number of runs for that inning by 0.23.

If you average the 36 values of the numbers on a player's card, you obtain the average amount by which he changes your expected number of runs scored in an inning in a single roll. This is not quite the same as saying that it is the result of a plate appearance, because occasionally the result of a roll will keep the batter at the plate. For example, it might be "Strike. Runner out stealing" or "Ball. W–base on balls." The latter means that if the pitcher has a W, which stands for Wild, the batter walks; otherwise you roll again.

For example, Hank Aaron's 1962 card has two 1's (a home run), a 5 and a 6 (extra-base hits of varying amount), eight numbers that are often a single, depending on the grade of the opposing pitcher, four 14's (usually a walk), a 16 (which is often "first on error" depending on the opposing team's center fielder), a 40 (which varies but is often an out), and eighteen numbers that are always outs. The total of the values of his 36 numbers is 3.45, and so the average is .096. If you rolled 670 times for him over the course of a season (roughly corresponding to his 660 plate appearances in 1962), he would increase the team's expected number of runs by $670 \cdot .096 = 64$. Sabermetricians (aficionados of extremely sophisticated baseball statistics) have a statistic called Batting Runs Above Average (BRAA), which tells how many runs a player increased his team's number of runs during the season compared to the result of an average player. Hank Aaron in 1962 had 58 BRAA, according to [8]. This suggests that the APBA card makers, my analysis, and the sabermetricians are pretty much in synch.

One of the uses that can be made of having these valuations of players' cards is to equalize fantasy teams. To do this, you need to be able to have a method for telling the value of pitching ratings, fielding ratings, and speed ratings on a basis comparable to the batting values. This is accomplished in Section 4.

Three strategy aspects which are evaluated are

- When should you Hit and Run? This is an option with a runner on first or runners on first and third.
- When should you "play it safe" with a slow runner on base? There are numbers which say things like "Single, runner to third, S out at third." So, if your runner on first has an S (for slow), you have an option to play it safe on a single. If you play it safe, the runner only goes to second on a single, regardless

of whether he would have ordinarily have gone to third or been out at third due to his S. My analysis tells when you should play it safe on a single, double, or fly ball.

- How should you align your outfielders? Each outfielder has a fielding rating, 1, 2, or 3. We determine, for each combination of these numbers, which alignment into left field, center field, and right field produces optimal results.

The analysis, although totally mathematical once it gets going, depends on input parameters taken from a batch of cards. I use a sample of 350 cards from the period 1956 to 1966, which is when I was actively playing the game. The parameters include what fraction of the time each batting number appears on all the cards, what fraction of the pitchers have each grade, A, B, C, and D, how often pitching adornments for strikeouts and walks occur, what fraction of the batters are fast (F) or slow (S), and what fraction of the fielders have the various fielding ratings. These fractions have a great effect on the values of the batting numbers. For example, the number 9 is often a single, but is usually an out against a grade A or C pitcher. Since my parameters say that the pitcher will have grade C 46% of the time, this makes 9 have a relatively poor rating, whereas if there were fewer C pitchers, it would have a better rating.

A very different version of this paper was written for APBA players who may not know any advanced mathematics.([3]) The purpose of the present paper is to explain how matrix methods can be applied to perform an interesting analysis of a game. Although Markov chains are present, no advanced theorems of Markov theory are involved. It is really just matrix manipulation and matrix equations.

Most of the ideas in this analysis were developed by the author in 1964, and a preliminary version of this analysis was performed then. It was limited by inadequate computer access at that time. The computer program used for the current analysis was the computer algebra system `Maple`. The new analysis was inspired by the *New York Times* article.

2. Details of evaluation of batting cards

We will work with various 25-by-25 transition matrices for the 25 (base,out) states. The possible bases occupied are 0, 1, 2, 3, 1-2, 1-3, 2-3, and 1-2-3. With none (resp., one, two) out, these comprise states 1 to 8 (resp. 9 to 16, 17 to 24). State 25 is the terminal state of three out. We write bases occupied as, for example, 1-3 to distinguish "runners on first and third" from situation number 13 (which is runners on first and second and one out) and from the strikeout number 13 on a batter's card.

For each integer k from 1 to 41, corresponding to the numbers that may appear on a batter's card[1], we form the 25-by-25 transition matrix M_k with (i, j)-entry $M_k(i, j)$ the probability of going to state j if you are in state i and k is rolled. (I find it convenient to use the inaccurate term that the number (from 1 to 41) is "rolled." The dice are rolled and then the number corresponding to the dice is found on the batter's card; it is a consequence of the roll and the batter's card, but I will say it is "rolled.") The rows of each M_k sum to 1. The last row of M_k is $[0, \dots, 0, 1]$. It is included mainly just so the theory of absorbing Markov chains applies. See, for example [4, §11.2]. The matrices M_k incorporate the various outcomes that can occur, depending upon the distribution of pitching and fielding ratings. They also incorporate the strategies of playing it safe and hit-and-running.

We illustrate with one example of a portion of an M_k. If 8 is rolled with a runner on first, it is an out, runner to second, against an A or B pitcher, while against a C or D pitcher it is single, runner to third, S out at third. We determine by methods discussed in Section 3 that, on average, an S runner on first should play it safe on a single against a C or D pitcher. Thus

$$M_8(i,j) = \begin{cases} A+B & \text{if } (i,j) = (2,11),\ (10,19), \text{ or } (18,25) \\ (C+D)S & \text{if } (i,j) = (2,5),\ (10,13), \text{ or } (18,21) \\ (C+D)(1-S) & \text{if } (i,j) = (2,6),\ (10,14), \text{ or } (18,22) \\ 0 & \text{if } i \in \{2,10,18\},\ j \text{ not as above.} \end{cases}$$

Here, for $t = 0$, 1, or 2, state $2 + 8t$ refers to runner on first and t out, while state $5 + 8t$ (resp. $6 + 8t$) refers to runners on first and second (resp. first and third) and t out. Also, A, B, C, and D are the fractions of the time that the pitcher will have each grade. For example $C = .46$. Also, S is the fraction of all cards that are S (Slow).

Then define a 25-by-25 transition matrix M_0 by

$$M_0 := \sum_{k=1}^{41} p_k M_k, \tag{2.1}$$

where p_k is the fraction of the time that the number k occurs on our batters' cards. The matrix M_0 is the transition matrix for an average batter. It is a Markov chain with one absorbing state, state 25. Let $\widetilde{M}$ denote the 24-by-24 matrix obtained from M_0 by deleting the last

[1]The actual program goes to 43 because two of the extra-base-hit numbers have an alternate interpretation, but we shall ignore this minor deviation in this paper.

row and column. Most of its rows do not sum to 0 because transitions to a third out are not included in $\widetilde{M}$. It is the transient submatrix associated to the absorbing Markov chain M_0. Most of our work will involve $\widetilde{M}$.

In linear algebra, it is common to have the transition probabilities appear in the columns rather than the rows. You can do it either way, but it seems more common in probability texts to put the transition probabilities in rows. The two ways would be equally convenient for our analysis.

First we compute the fraction of the time that a batter is in each of the 24 states. Let $\mathbf{q} = [q_1, \dots, q_{24}]$ be a row vector of probabilities of being in the various states prior to a roll, with $\sum_{i=1}^{24} q_i \le 1$. This sum might be less than 1 because it does not include the probability that the inning has already ended. Then $\mathbf{q}\widetilde{M}$ is the row vector of probabilities of being in the various states after the roll, and the sum of its entries may be even smaller than it is for $\mathbf{q}$. Let $\mathbf{e} = [1, 0 \dots, 0]$, a row vector of length 24. It represents the state at the beginning of the inning. Then

$$\mathbf{p} := \mathbf{e} + \mathbf{e}\widetilde{M} + \mathbf{e}\widetilde{M}^2 + \mathbf{e}\widetilde{M}^3 + \cdots \tag{2.2}$$

has as its ith entry the expected number of times that situation i will occur during an inning. The general theory of Markov chains with an absorbing state implies that this infinite series converges to a finite row vector. Let $\mathbf{pr}$ denote the row vector obtained by dividing each entry of $\mathbf{p}$ by the sum of the entries of $\mathbf{p}$. Then $\mathbf{pr}$ gives the probabilities $\mathrm{pr}(i)$ of being in each of the 24 states.

The easiest way to compute $\mathbf{p}$ is derived by first multiplying (2.2) on the right by $\widetilde{M}$, obtaining

$$\mathbf{p}\widetilde{M} = \mathbf{e}\widetilde{M} + \mathbf{e}\widetilde{M}^2 + \mathbf{e}\widetilde{M}^3 + \cdots .$$

Combining this with (2.2) yields that $\mathbf{p}\widetilde{M} = \mathbf{p} - \mathbf{e}$, and so

$$\mathbf{p}(I - \widetilde{M}) = \mathbf{e}, \tag{2.3}$$

where I is the 24-by-24 identity matrix. The matrix equation (2.3) can be solved for $\mathbf{p}$. It is perhaps more commonly thought of as the transposed version, $(I - \widetilde{M})^T \mathbf{p}^T = \mathbf{e}^T$. In this latter version, you are working with column vectors, and, if you are thinking of it as a system of 24 linear equations in 24 unknowns, the equations line up more naturally. At any rate, it is by solving this system of equations, or matrix equation, that $\mathbf{p}$ and then $\mathbf{pr}$ is obtained. This is easily done by `Maple`, yielding that $\mathbf{pr}$ is as in Table 1. A more explicit formula for $\mathbf{p}$ is that it equals the first row of $(I - \widetilde{M})^{-1}$ or the first column of $(I - \widetilde{M^T})^{-1}$, but it is easier just to solve (2.3).

TABLE 1. Fraction of the time a batter is in each situation

		outs		
		0	1	2
	0	.24239	.17284	.13518
	1	.06248	.07462	.07359
Bases	2	.01671	.03318	.04347
	3	.00189	.00728	.01491
	1-2	.00924	.01664	.01895
	1-3	.00584	.01128	.01744
	2-3	.00496	.01013	.01305
	1-2-3	.00233	.00515	.00645

Now we explain how we found the expected number of runs scored in an inning subsequent to being in situation j, which we denote by $E(j)$. Let $\mathbf{E}$ be the column vector of length 24 with entries $E(j)$, and let $E(25) = 0$. If you are in situation i and roll a k, the number of runs that you expect to score in the remainder of the inning, including on that roll, is

$$\sum_{j=1}^{25} M_k(i,j)(E(j) + r_{k,i,j}),$$

where $r_{k,i,j}$ is the number of runs scored on that roll (rolling k and going from state i to j). Then

$$E(i) = \sum_{k=1}^{41} p_k \sum_{j=1}^{25} M_k(i,j)(E(j) + r_{k,i,j}), \tag{2.4}$$

or equivalently

$$\mathbf{E} = \widetilde{M}\mathbf{E} + \mathbf{b}, \tag{2.5}$$

where $\mathbf{b}$ is a column vector of length 24 whose ith entry is

$$\sum_{k=1}^{41} p_k \sum_{j=1}^{25} M_k(i,j) r_{k,i,j}.$$

This (2.4) is the key equation, expressing the desired E-values in terms of other E-values, including itself, and the r-values. Once the r-values are known, it can be solved to find the important E-values.

Most of the time, $r_{k,i,j}$ does not depend on k and equals

$$r(i,j) := 1 + BR(i) + \left[\tfrac{i-1}{8}\right] - BR(j) - \left[\tfrac{j-1}{8}\right], \tag{2.6}$$

where

$$BR(i) := \begin{cases} 0 & i \equiv 1 \pmod 8 \\ 1 & i \equiv 2,3,4 \pmod 8) \\ 2 & i \equiv 5,6,7 \pmod 8 \\ 3 & i \equiv 0 \pmod 8 \end{cases}$$

is the number of base runners in situation i, and $[\frac{i-1}{8}]$, which denotes the integer part of the fraction, is the number of outs in situation i. The 1 in (2.6) is due to the batter. The formula (2.6) does not work when $j = 25$, when $r_{k,i,j}$ is usually 0. Let $\mathbf{r}$ be a column vector of length 24 whose ith entry is $\sum_{j=1}^{24} \widetilde{M}(i,j) r(i,j)$. This gives, except for the exceptional situations noted below, the average number of runs scored on a single roll from situation i. Note that $r(i,j)$ is sometimes negative, which is nonsensical, but this will never happen if $\widetilde{M}(i,j) \neq 0$.

There are two types of exceptions when the formulas of the preceding paragraph do not give the number of runs scored on the play. Those with $j = 25$ are the rare cases, such as 9 with a runner on third and two out against a B or D pitcher[2], in which runs score on the play, but then a base runner makes the third out. Let $d(k,i)$ denote the expected number of runs which are scored on inning-terminating plays when rolling a k from situation i. For example, $d(9,20) = B + D$, since situation 20 is (base,out)=(3,2). Let $\mathbf{d}$ be a column vector of length 24 whose ith entry is $d_i := \sum_{k=1}^{41} p_k d(k,i)$, the average number, out of all rolls from situation i, of runs scored on plays which include the third out of the inning. These numbers d_i will be 0 unless $i \geq 20$.

A bigger exception is the cases in which the batter stays up after the roll, such as rolling a 37 with the bases empty[3] if the pitcher does not have a W. We have $M_{37}(1,1) = 1 - W$, where W is the fraction of all pitchers that have a W. The formula (2.6) would incorrectly say that one runner scored on the play. We define a column vector $\mathbf{No}$ of length 24 which has ith entry $\text{No}_i := \sum_{k=1}^{41} p_k U(k,i)$, where $U(k,i)$ is the probability that the batter stays up if you roll k in situation i. The

[2]single, runner scores, batter out trying for second

[3]Strike. W-Base on balls.

above example for 37 says $U(37,1) = 1-W$ and contributes $p_{37}(1-W)$ to No_1.

Now we can write $\mathbf{b} = \mathbf{r} + \mathbf{d} - \mathbf{No}$, and (2.5) becomes

$$(I - \widetilde{M})\mathbf{E} = \mathbf{r} + \mathbf{d} - \mathbf{No}. \tag{2.7}$$

This equation was solved for $\mathbf{E}$ by `Maple` to obtain Table 2.

TABLE 2. Expected number of runs from different situations

		outs		
		0	1	2
	0	0.4327	0.2254	0.0792
	1	0.8071	0.4737	0.1892
Bases	2	0.9833	0.6055	0.2862
	3	1.2053	0.8693	0.3481
	1-2	1.3294	0.8288	0.4042
	1-3	1.6274	1.0519	0.4450
	2-3	1.7479	1.1924	0.5279
	1-2-3	2.1726	1.4705	0.7098

The most interesting value in $\mathbf{E}$ is the expected number, $E(1)$, of runs scored when no one is on and no one out, since that tells the average number of runs scored in an inning. This value, 0.4327, after being multiplied by 9, yields the average number of runs scored by a team in a 9-inning game. This value, 3.8943, is quite consistent with actual baseball figures during the early 1960's on which this analysis is based. It would be higher if we based our analysis on the cards from the late 1990's.

The values $V(k)$ of the numbers on a card are now easily obtained as

$$\begin{aligned} V(k) &= \sum_{i=1}^{24} \mathrm{pr}(i)\Bigg(\sum_{j=1}^{24} M_k(i,j)\big(r(i,j) - U(k,i) + E(j) - E(i)\big) \\ &\qquad + M_k(i,25)\big(d(k,i) - E(i)\big)\Bigg). \end{aligned}$$

This is obtained by averaging, over all initial situations i and all subsequent situations j obtained when rolling a k, the number of runs obtained on that roll plus the change in expected number of subsequent runs to be obtained later in the inning. This yields numbers such as $V(1) = 1.4101$ and $V(13) = -0.2317$, which were mentioned in Section 1. They are listed in [3], but we will not list them here, as they would only be of interest to an APBA player.

3. Strategy

In this section, we discuss briefly the way in which the results of Tables 1 and 2 are used in making strategy decisions of the three types outlined in Section 1. The strategy decisions also affect the computation of the numbers in those tables, so it is sort of a feedback loop. The way this was performed was to first make an initial assumption about hit-and-running, playing-it-safe, and aligning your outfielders, then use that assumption to compute numbers $\mathrm{pr}(i)$ and $E(i)$ for the two tables, next use these values to determine new strategies, and then finally recompute the numbers $\mathrm{pr}(i)$ and $E(i)$ using the new strategy. After this, the process stabilizes because the strategies based on the revised tables are the same as the strategies obtained from the earlier tables.

The biggest difference when hit-and-running as compared to an ordinary at-bat is that 13, instead of being a strikeout, is "runner out stealing; if runner has an 11 on card, he steals safely." Some other numbers change, too, but this is the most significant. Because of this, one's intuition is to hit and run if the runner has an 11, and not if he doesn't. The numbers in Table 2 can be used to verify that this is the best strategy. We compare, for the situations in which hit-and-running is allowable, namely $i = 2, 6, 10, 14, 18$, or 22, the value of

$$\sum_{k=1}^{41} p_k \sum_{j=1}^{25} M_k(i,j)(E(j) + r_{k,i,j}) \tag{3.1}$$

if $M_k(i,*)$ is the transition matrix[4] for (a) hitting away, (b) hit-and-running with an 11 on first, and (c) hit-and-running without an 11 on first. We obtain the largest value when hit-and-running with an 11, and the smallest value when hit-and-running without an 11.[5] This is all based on an average batter. For a specific batter, your strategy might be different, but you can still use Table 2 to help you make your decision.

A similar analysis is performed to decide whether to play it safe if an S-runner is on base in a situation in which "S is out" as in the earlier discussion of 8 with a runner on first. For each situation i, one compares (3.1) using two different $M_k(i,*)$'s and the associated $r_{k,i,j}$'s, one with

[4] and $r_{k,i,j}$ the runs scored on that roll

[5] If there is just a runner on first and he has a 10 but not an 11, because of a different stolen-base situation which is advantageous if the runner has a 10 or an 11, hit-and-running turns out to be a "wash" if the runner has a 10 but not an 11. Largely for programming convenience, we decided not to hit and run in this situation.

playing it safe and one not playing it safe. In [3], all the conclusions are listed, but they would not be of much interest to anyone except an APBA player. Some of them are quite delicate. For example, with runners on second and third, with an S runner on second, always play it safe on a single against an A pitcher, never play it safe against a B pitcher, and against a C or D pitcher play it safe with less than two outs.

Decisions about playing it safe on a fly ball with an S runner on third and less than two out are even more complicated because they depend both on the pitcher (since some hit numbers are "fly out, runner scores" against certain types of pitchers) and all the outfielders, since whether certain out numbers are "fly out, runner scores," "fly out, runner holds," or "fly out, runner scores, S out at home" depend on the fielding column of specific outfielders. Here again, in practice, an APBA player could use Table 2 together with information about the specific batter, pitcher, and fielders to make a decision. But for determining the final transition matrices for the "average" player, we needed to include specific rules about when we were playing it safe, and those are listed in [3].

Deciding how to align your outfielders was handled differently. We determine this by running the `Maple` program with each of the various possible combinations of fielding numbers, and seeing which alignment gives the lowest expected number of runs scored. An outfielder's fielding rating is 1 (the worst), 2, or 3 (the best). The program has parameters WLF (worst left fielder), MLF (middle left fielder), and BLF (best left fielder), and similarly for center fielders and right fielders, telling the probability that the particular fielder has a certain fielding rating. Ultimately these will be given values which are the fraction of the time that the left fielder is of each of these fielding types, etc., and this depends, not only on the cards from which we are sampling, but also on the alignment strategy.

Here is an example of determining alignment strategy. If you have two 2's and a 3 in the outfield, we run the program first with BLF=1, MCF=1, and MRF=1, and the other probabilities, such as MLF, equal to 0. Thus this run of the program assumes that you have put your 3 in left field. The program determines that the expected number of runs in an inning is .42463 if the 3 is in left, .43322 if the 3 is in center, and .42234 if the 3 is in right, so you should put your 3 in right field to minimize the opponent's expected number of runs. A similar analysis is done for each combination of fielding numbers.

4. Values for speed, pitching and fielding

Now all the strategies determined in the preceding section are implemented into the `Maple` program. For hit-and-running, we sample that 9% of the cards have an 11, and we estimate that you will hit and run 80% of the time when you have a runner with an 11 on first, so we say that you are hit and running 7% of the time when a runner is on first or first and third. Thus the transition matrices in the rows corresponding to these base situations have the usual outcome weighted by $1 - H$ and the hit-and-run outcome weighted by H, where H is a parameter which is set to .07. The values of $E(i)$ listed in Table 2 are contingent on all these strategies.

To determine the value of an S (slow base runner), we compute

$$\sum_{i=1}^{24} \mathrm{pr}(i) \sum_{k=1}^{41} p_k \sum_{j=1}^{25} M_k(i,j)(E(j) + r_{k,i,j}) \tag{4.1}$$

with M_k (and the associated $r_{k,i,j}$) based on having $S = 1$, and then again with $S = 0$, and take the difference. This will give the change in expected runs scored in an average roll by having an S runner on an affected base. This equals $-.01898$. But to compare this with a batting number, there are several considerations.

One is that the player with the S is not going to always be on base. The more frequently the player is on base, the more disadvantageous his S rating is. But my analysis cannot measure such a fine distinction. We must assume that each player on the team is equally likely to be on base. Suppose a team has, in its lineup, two S players. Then two ninths of the time a specific baserunner (such as the runner on second) would have an S. So the average loss to the team on any roll due to the S's would be $\frac{2}{9} \cdot .01898$. But a given player with an S will only be causing half of this loss. Thus the average loss caused by the player's S is .01898/9.

But this analysis is happening every play of the game (while your team is at bat). An S runner could be on an affected base while several batters are up. The previous paragraph takes this into account. If you average the values of the 36 numbers on a batter's card, this gives the average amount by which he increases the team's expected number of runs on a single roll. On average, a batter will be up 4.5 times per game, and so the average of the values of the numbers on his card should be multiplied by 4.5 to give the amount by which his batting numbers increase the team's expected number of runs during a game. The value .01898/9 that a person's S hurts you on every roll of the game should be multiplied by 40.5, for the 40.5 rolls during a game, on average. Since 40.5/9 equals 4.5, over the course of a game the .01898 negative value of an S is exactly comparable to the average value of the numbers on the players card. If comparing it with a single number on the player's card, its .01898 should be multiplied by 36, yielding .683, since the sum of the values of the numbers on the card had to be divided by 36 to form the average. Having an S turns out to be roughly equal to the difference between one of your 36 batting numbers being a 7 (usually a single, although occasionally an out against a good pitcher) rather than a pure out number (and all other numbers unchanged), since $V(7) = .47$ and $V(13) = -.23$, and $.47 - (-.23)$ is approximately equal to .683.

A similar analysis can be made for F and the base-running value of an 11.

To find the value of an A pitcher, one merely needs to run the program with the pitcher parameters set as $A = 1$, $B = 0$, $C = 0$, $D = 0$, and find that then the expected number of runs in an inning is .25186, which is .18087 less than the value of .43273 against an average pitcher. Thus the A pitcher saves .18087 runs per inning. But how do we compare this with the value of a batter's numbers? It should be done on a per-game basis. A batter bats 4.5 times per game, and so we multiply the average of the values of his numbers by 4.5 to see how many runs per game do his plate appearances help the team's expected number of runs. The A pitcher pitches roughly 7 innings every fifth game, hence 1.4 innings per game (of his team). Thus the A pitcher's value, per game of his team, is $.1898 \cdot 1.4 = .2532$. If he also has a Z, for good control, that adds another $.04 \cdot 1.4 = .056$, bringing his value to .31. If this is multiplied by 162, the number of games in a season, this brings his value to the team to about 50 runs during the season, in rough agreement with the sabermetricians' value for the Runs Above Average Pitcher([8]) of a very good pitcher, again establishing a nice compatibility among the APBA card makers, my analysis, and sabermetrics.

If we wish to evaluate a batter's card by the sum of the values of its batting numbers, under the system in which 1 (a home run number) has 1.41, 13 (a strikeout number) $-.23$, and Hank Aaron a total of 3.45, then the value of an A pitcher with a Z would be $.31 \cdot 36/4.5$ (because the sum of the batting numbers didn't take into account the batter's 4.5 at bats per game). So the A pitcher with a Z is worth 2.48 to the team if Hank Aaron's batting is worth 3.44. Aaron's speed and fielding would make him even more valuable.

A similar analysis is made for the other pitching grades and also for fielding. It turns out that the difference between the best fielding column and the worst fielding column, on a scale such as that of the previous paragraph comparable to the sum of the values of the batting numbers, is approximately 1 at each position. Tables for all these appear in [3], as does an annotated version of the `Maple` program which was used to perform the calculations.

In conclusion, the analysis described above, accompanied by the detailed results in [3], should enable APBA players to equalize or optimize their fantasy teams and to make informed decisions about strategy. The method of analysis employed here provides a paradigm for analyzing games of a certain type.

References

[1] APBA–Between the Lines, `forums.delphiforums.com/apbabtl`.

[2] B. Bukiet, E.R. Harold, and J. Palacios, *A Markov chain approach to baseball*, Operations Research **45** (1997) 14-23.

[3] D.M. Davis, *Evaluating APBA cards*, `www.lehigh.edu/~dmd1/apba1.pdf` (2009).

[4] C.M. Grinstead and J.L. Snell, *Introduction to probability*, Amer Math Soc (1997). Also at `math.dartmouth.edu/~prob/prob/prob.pdf`.

[5] C. Hine, *APBA lets stat lovers be managers*, New York Times, August 9, 2009.

[6] M.D. Pankin, *Baseball as a Markov chain*, `www.pankin.com/markov/intro.html`

[7] J.S. Sokol, *An intuitive Markov chain lesson from baseball*, Informs Transactions on Education **5** (2004) 47-55.

[8] J. Thorn, *Total Baseball*, Warner (2004).

Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA

E-mail address: `dmd1@lehigh.edu`