An Introduction to Markov Chains and Baseball

Ethan Greenly

December 2, 2020

Abstract

Baseball has been entrenched in America's history for more than a century. From its early beginnings in Cooperstown after the Civil War to Jackie Robinson breaking the color barrier to modern day baseball, there is a reason why baseball is called "America's Pastime."[3] Those who passionately follow baseball will know about all of the statistics like RBIs, ERAs, batting averages, etc.. These statistics have long been around to tell what happens in the game. But up until recently, faith in these statistics as an accurate measure of a player's worth has never been challenged. After sabermetrics, a form of advanced statistical analysis, proved to be successful for the Oakland Athletics, chronicled in the movie *Moneyball*, more teams began to use sabermetrics in order to better understand if a player was a high performer or would fit into their strategy. This has led to many different new statistics such as WAR (Wins Above Replacement) or wOBA (Weighted On-Base Average). But what mathematics do these statistics rely upon? One answer to this is Markov chains. Markov chains are a beautiful way to look at baseball and figure out the likelihood of entering the various states of the game and how they progress and are used to calculate the expected number of runs beginning at a given state in a given inning. Throughout this paper, we will study the properties of Markov chains and how they can accurately describe some baseball statistics.

1 Introduction

A.A. Markov was a student of Chebyshev at St. Petersburg University, eventually teaching the probability courses upon Chebyshev's resignation in 1883.[6] He developed his theory of Markov chains in 1913 after delving into a poem entitled "Eugene Onegin" by Alexander Pushkin. By looking at the patterns of vowels and consonants, Markov developed a new technique for studying certain stochastic processes (now called Markov chains) and moved the theory of probability in this new direction, one with chains of linked events. Markov chains are now found in many different fields of academics, from physics to biology to computer science, e.g., helping researchers identify genes in DNA and powering algorithms in voice search and web search.[4]

2 Properties

Stochastic processes are processes with chance involved, indexed against some other variable or set of variables, such as time.[5] Markov chains are an important example of stochastic processes. They are mathematical processes that experience a transition from one state to another according to probabilistic rules. What sets Markov chains apart from normal stochastic processes is that the probability of a future state depends only upon the present state and how much time has passed. There is no need to keep a record of the states that were occupied to get to the present state. This "memory-less" property is called the *Markov property*.[2]

Let us say that we were in the world of the movie *Groundhog Day* (starring Bill Murray). He gets stuck in a never ending loop of the same day, which happens to be Groundhog Day. Let us assume that the probability of him repeating the day or moving onto the next day is only determined by the present state he is in, where each state might represent his behavior or motives in that given day. This is a prime example of a Markov property. No matter what previous states he has been in, the only thing that determines the probability of the next state is the current state he is in. The mathematical way of stating this for a Markov chain is $\{X_n | n = 0, 1, 2, 3, ...\}$, is that $P(X_n = i_n | X_{n-1} = i_{n-1}) = P(X_n = i_n | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1})$.

In order to record all of these probabilities of moving from one state to another in our Markov chain, a transition matrix can be created. The $(i, j)^{th}$ element of the transition matrix $\mathbb{P}$, denoted $P_{i,j}$, is equal to the probability that the random variable X_t moves from state i to state j in one step, or written mathematically, $P_{i,j} = P(X_{t+1} = j | X_t = i)$. Because each element of the matrix is a probability, this element $P_{i,j}$ is greater than or equal to zero and less than or equal to 1. Thus each row of the matrix is a probability vector, with a sum of one.[7]

3 Proofs

Let us first note that the Markov process $\{X_n | n = 0, 1, 2, 3, ...\}$ is completely determined once the following theorem is established and the probability distribution of X_0 is specified.

Theorem 1: Suppose $P(X_0 = i) = P_i, i = 0, 1, 2, 3, \ldots$, and $P(X_{n+1} = j | X_n = i)$

$= P_{ij} \; \forall \, i, j = 0, 1, 2, 3, \ldots$

Then $P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \cdots P_{i_0, i_1} P_{i_0}$.

Proof. We proceed by induction on $n \in \mathbb{Z}^+$. Let $n = 1$. Then $Pr(X_0 = i_0, X_1 = i_1) = Pr(X_1 = i_1|X_0 = i_0)Pr(X_0 = i_0) = P_{i_0,i_1}P_{i_0}$ (by the definition of conditional probability). This completes the trivial case.

Let $n \in \mathbb{Z}^+$ and suppose $Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P_{i_{n-1},i_n} \cdots P_{i_0,i_1}P_{i_0}$.
Then $Pr(X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}) =$

$Pr(X_{n+1} = i_{n+1}|X_0 = i_0, \ldots, X_n = i_n) \cdot Pr(X_0 = i_0, \ldots, X_n = i_n)$ (by the definition of conditional probability).

By using the definition of Markov process and applying the induction hypothesis, the above equation becomes $Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n, X_{n+1} = i_{n+1}) =$

$Pr(X_{n+1} = i_{n+1}|X_n = i_n) \cdot P_{i_{n-1},i_n} \cdots P_{i_0,i_1}P_{i_0}$. Since we know that

$Pr(X_{n+1} = i_{n+1}|X_n = i_n) = P_{i_n,i_{n+1}}$, the above equation reduces to

$Pr(X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}) =$

$Pr(X_{n+1} = i_{n+1}|X_n = i_n) \cdot P_{i_{n-1},i_n} \cdots P_{i_0,i_1}P_{i_0} = P_{i_n,i_{n+1}}P_{i_{n-1},i_n} \cdots P_{i_0,i_1}P_{i_0}$.

This completes the induction and the proof. □

Andrey Kolmogorov was a Russian mathematician who laid the foundations for the modern-day study of Markov processes.[6] He and Sydney Chapman, a British mathematician, independently derived the following property of transition matrices that are used with Markov chains.

Theorem 2 (Chapman-Kolmogorov Equation): The transition matrix of n steps, $\mathbb{P}^n$, is equal to the transition matrix of r steps, $\mathbb{P}^r$, multiplied by the transition matrix of s steps, $\mathbb{P}^s$, where $n = r + s$. Stated simply $\mathbb{P}^n = \mathbb{P}^r\mathbb{P}^s$.

Proof. The law of total probability states that if A is an event and if $B_1, B_2, B_3, ...$ are mutually exclusive events, then $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + \cdots$. Thus $P_{ij}^{(n)}$, the probability of moving from state i to state j in n steps, is given by $P_{ij}^{(n)} = P(X_n = j | X_0 = i) = P((\bigcup_{k=0}^{\infty} X_n = j, X_r = k) | X_0 = i)$. By the Law of Total Probability, the latter probability is equal to $\sum_{k=0}^{\infty} P(X_n = j, X_r = k | X_0 = i)$. By the definition of conditional probability, this sum is equal to

$\sum_{k=0}^{\infty} P(X_n = j | X_r = k, X_0 = i) \cdot P(X_r = k | X_0 = i)$.

Applying the Markov property, terms of this sum yield

$P_{ij}^{(n)} = \sum_{k=0}^{\infty} P(X_n = j | X_r = k) \cdot P(X_r = k | X_0 = i) = \sum_{k=0}^{\infty} P_{ik}^{(r)} P_{kj}^{(s)}$, since $s = n - r$.

Hence $\mathbb{P}^n = \mathbb{P}^r \mathbb{P}^s$! □

Corollary 2: A corollary of the above theorem is that the n-step transition matrix, $\mathbb{P}^n$, is equal to the n-th power of the one-step transition matrix. In mathematical terms, $\mathbb{P}^n = P^n$.

Proof. We will proceed by induction on $n \in \mathbb{Z}^+$. If $n = 1, \mathbb{P}^1 = P^1 = P$, the one-step transition matrix.

Let $n \in \mathbb{Z}^+$ and assume $\mathbb{P}n = P^n$. Then $\mathbb{P}^{n+1} = \mathbb{P}^n \cdot \mathbb{P}^1 = P^n \cdot P = P^{n+1}$ (by the Chapman-Komolgorov Equation). This completes the induction and the proof.

Thus $\mathbb{P}^n = P^n$. □

An absorbing Markov chain is one in which for some states it is impossible to leave that state, and the process that is in any state could, after a number of steps, reach that aforementioned state. An absorbing state is defined as one in which $P(X_{t+1} = i)|X_t = i) = 1$. A Markov chain must not just have absorbing states in order to be considered an absorbing Markov chain. It must also have the property that from all other states, it eventually reaches an absorbing state with a probability of one.[8] Let us now prove this fact.

Theorem 3: In an absorbing Markov chain, the probability that the process will be absorbed is 1.[9] (i.e., $lim_{n\to\infty}\mathbb{P}^n = 0$ meaning if $\mathbb{P}^n = [P_{ij}^n]$, then $lim_{n\to\infty}P_{jk}^n = 0\forall j,k$)

Proof. Let $\{X_n | n = 0, 1, 2, 3, ...\}$ be an absorbing Markov chain. Thus from each state s_j it is possible to reach an absorbing state. Let m_j be the minimum number of steps to reach an absorbing state starting from state s_j. Since circuitous routes to an absorbing state are possible, let P_j be the probability that it will take more than m_j steps to reach an absorbing state starting from state s_j. Because there are multiple ways to reach the absorbing state starting from state s_j, $P_j > 0$. Because there is still a possibility of reaching an absorbing state in m_j steps, $P_j < 1$. If there are k states in the Markov chain, let $M = max\{M_1, M_2, \ldots, M_k\}$ and let $P = max\{P_1, P_2, \ldots, P_k\}$. The probability that the process starting in state s_j is not absorbed in M steps is less than or equal to P. The probability that it is not absorbed in $2M$ steps is the probability of it not being absorbed in M steps and it not being absorbed in another M steps, which is less than or equal to P^2. Because $P < 1$, continuing on this track, these probabilities, P^n, approach 0. Thus, because the probability of not being absorbed in n steps is monotone decreasing and converges to 0, $lim_{n\to\infty}P_{jk}^n = 0$, so

$lim_{n\to\infty}\mathbb{P}^n = 0$. In conclusion, the probability that an absorbing Markov chain will reach an absorbing state (be absorbed) is equal to 1. $\square$

With absorbing Markov chains, there is also a neat way to write the infinite sum of the powers of these transition matrices. This will prove to be helpful later when we get to the applications of Markov chains to baseball.

Theorem 4: Let $\mathbb{T}$ be the nxn transition matrix associated with an absorbing Markov chain. The matrix $I - \mathbb{T}$ has an inverse Q, where $Q = I + \mathbb{T} + \mathbb{T}^2 + \ldots$.[9]

Proof. Suppose $(I - \mathbb{T})\mathbf{x} = 0$. This is equivalent to $\mathbf{x} = \mathbb{T}\mathbf{x}$. Multiplying both sides of this equation by $\mathbb{T}$ on the left, we get $\mathbb{T}\mathbf{x} = \mathbb{T}^2\mathbf{x}$; since $\mathbf{x} = \mathbb{T}\mathbf{x}$, then $\mathbf{x} = \mathbb{T}^2\mathbf{x}$. Repeating this process n times, we get $\mathbf{x} = \mathbb{T}^n\mathbf{x}$. Using Theorem 3 where we proved that $lim_{n\to\infty}\mathbb{T}^n = 0$, we can take the limit of both sides of this equation to find that $\mathbf{x} = 0$. Since $(I - \mathbb{T})\mathbf{x} = 0$ has only the trivial solution, $I - \mathbb{T}$ is invertible.

Let $\mathbb{Q} = (I - \mathbb{T})^{-1}$.

Notice that the following equation holds, $(I - \mathbb{T})(I + \mathbb{T} + \mathbb{T}^2 + \cdots + \mathbb{T}^n) = I - \mathbb{T}^{n+1}$.

If we multiply both sides of this equation by the matrix $\mathbb{Q}$ on the left, we get

$I + \mathbb{T} + \mathbb{T}^2 + \cdots + \mathbb{T}^n = \mathbb{Q}(I - \mathbb{T}^{n+1})$. Taking the limit of both sides, we get $\mathbb{Q} = I + \mathbb{T} + \mathbb{T}^2 + \cdots$.

Therefore $(I - \mathbb{T})^{-1} = I + \mathbb{T} + \mathbb{T}^2 + \cdots$. $\square$

4 Applications to Baseball

Markov chains are useful for figuring out the number of expected runs from any state in a baseball game to the end of the inning. We think of baseball in terms of 28 different states. Twenty-four of these states relate to locations of runners on base with a specified number of outs, such as no runners on and no one out to bases loaded with two outs. There are 24 states because with three bases (we don't count home as a base here because there is always a batter up to the plate) and there can either be a runner on that base or not. Thus for each base there are two choices, so altogether $2^3 = 8$ options. Then there can be 0, 1, or 2 outs, thus a total of $8 * 3 = 24$ states. The last four states are absorbing states which are the 3-out states. You can either score no runs, one run, two runs, or three runs at the end of the inning (you cannot score four runs or there would be no outs on the play). Thus, by analyzing every play in which the batter changed, we can create a transition matrix for the game of baseball. But how do we get to the number of expected runs? We first have to create a runs matrix consisting of the number of runs scored at every state transition of the game. For example, going from bases loaded and 2 outs to no one on and two outs means that a grand slam has occurred, i.e. 4 runs were scored. The figure below gives a picture of such a matrix, showing how many runs are scored in each transition. The states are numbered increasingly by the number of men on base and the bases they occupy. The zero out states are from 1-8, 1 out states are from 9-16, the two out states are from 17-24, and the 3 out absorbing states are 25-28.

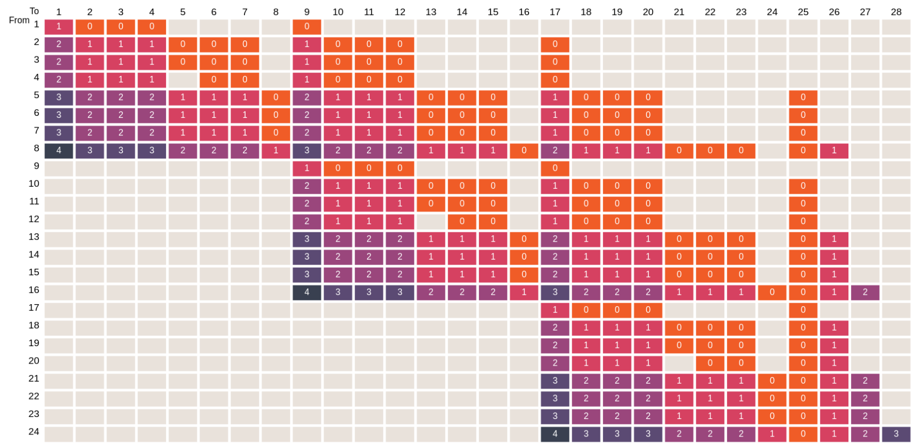

Figure 1: Number of runs scored in transition from each state[11]

This matrix can be used to create an expected runs column vector R, whose i^{th} entry gives the expected runs scored in one play starting in state i. We left multiply R by the transition matrix $\mathbb{T}$ that can be created from available game data to obtain the expected runs scored in the subsequent play. Now in order to get our expected runs vector E, we have to add each such (modified) expected runs vector $R, \mathbb{T}R, \mathbb{T}^2R$, etc. for all possible subsequent plays. In mathematical terms this would look like $E = R + \mathbb{T}R + \mathbb{T}^2R + ...$ Summing an infinite series of vectors would seem difficult. But with Theorem 4 we can make it much simpler. There are two ways we ensure that $I - \mathbb{T}$ is invertible. Either we make the transition matrix have all 1s in the appropriate places in the 25-28 by 25-28 block of matrix $\mathbb{T}$ or we can completely ignore those states because the team will not score any additional runs starting in these states. Either way, this will make sure that the $I - \mathbb{T}$ matrix is invertible. Thus, $E = R + \mathbb{T}R + \mathbb{T}^2R + ... = (I - \mathbb{T})^{-1}R$. Now we are able to examine the expected number of runs after a given state for the remainder of the inning. The figure below shows the number of expected runs from each state to the end of the inning for Major

League Baseball from 1921-2017.[10,11]

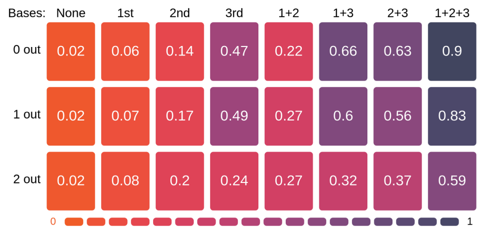

Figure 2: Expected Runs from 1921-2017[8]

Now what can we do with this information? We can create strategies for in-game play. Is it better to have a sacrifice bunt to move a baserunner forward, but creating an out in the process? Is it better to walk the batter or to pitch to him? We can answer this question more clearly by using a transition matrix for specific batters. This would not allow us to use the expected runs equation above that makes it easy to calculate E, but it would make it more accurate by using data for the specific batters. However these matrices can also be misleading. For example, I created an expected runs vector for Franscisco Lindor for the shortened 2020 MLB season. He has no expected runs for bases loaded and no outs because he was not up to bat in that situation in the shortened 2020 season.

	xxx	1xx	x2x	xx3	12x	1x3	x23	123
0 outs	.4106	.7301	.7449	1.7449	0	.4037	.375	0
1 out	.2254	.5365	.7278	.7790	.2917	.6537	.375	.375
2 outs	.1223	.3075	.3125	.4358	.375	.5	0	.5

Figure 3: Fransisco Lindor Expected Runs, 2020

5 Conclusion

Markov chains are a great example of a mathematical stochastic process used to compute probabilities where the probability of moving to the next state is only affected by the current state. We can use these to calculate the expected number of runs scored for the remainder of a given inning of a baseball game. Given this expected runs column vector, we are able to use it in a variety of ways to make a baseball team more competitive and give them an edge over their opponent.

6 References

1. Statshacker. The Markov Chain Model of Baseball. statshacker. statshacker, July 26, 2018. http://statshacker.com/blog/2018/05/07/the-markov-chain-model-of-baseball/.

2. Maltby, Henry, Worranat Pakornrat, and Jeremy Jackson. Markov Chains. Brilliant. Accessed November 6, 2020. https://brilliant.org/wiki/markov-chains/#: : text=A%20Markov%20chain%20is%20a,possible%20future%20states%20are%20fixed.

3. Baseball History, American History and You. Baseball Hall of Fame. Accessed November 18, 2020. https://baseballhall.org/baseball-history-american-history-and-you.

4. Hayes, Brian. First Links in the Markov Chain. American Scientist, September 6, 2018. https://www.americanscientist.org/article/first-links-in-the-markov-chain.

5. Stochastic Process. Encyclopdia Britannica. Encyclopdia Britannica, inc. Accessed October 26, 2020. https://www.britannica.com/science/stochastic-process.

6. O'Connor, J. J., and E. F. Robertson. Andrei Andreyevich Markov. MacTutor, August 2006. https://mathshistory.st-andrews.ac.uk/Biographies/Markov/.

7. Karlin, Samuel, and Howard M. Taylor. A First Course in Stochastic Processes. New York, NY: Acad. Press, 1975.

8. Maltby, Henry. Absorbing Markov Chains. Brilliant Math & Science Wiki. Accessed November 11, 2020. https://brilliant.org/wiki/absorbing-markov-chains/.

9. Thomay, Cal D. Markov Chain Theory with Applications to Baseball. The College Of Wooster Libraries, 2014. Markov Chain Theory with Applications to Baseball.

10. Pankin, Mark D. MARKOV CHAIN MODELS: THEORETICAL BACKGROUND. Markov Baseball Models Theory. Accessed November 11, 2020. http://www.pankin.com/markov/theory.htm.

11. Calestini, Lucas. The Elegance of Markov Chains in Baseball. Sports Analytics. Medium, September 10, 2018. https://medium.com/sports-analytics/the-elegance-of-markov-chains-in-baseball-f0e8e02e7ac4.