Collecting Baseball Cards with Jordan Normal Forms

Scott Atkinson, Jeannette Imbrogiano Brown, Neil Calkin, Tyler Lemburg

June 24, 2010

Abstract

In this paper, we consider the baseball card collector's problem. We find a closed-form expression for the probability of getting 2 complete sets of n cards taken 1 at a time with replacement after t selections. This can be found by examining the t-th power of a transition matrix A. We use the Jordan Normal Form of this transition matrix to compute this probability more easily.

1 Introduction

1.1 Baseball Card Collector's Problem

The combinatorial problem we will be considering in this paper is the baseball card collector's problem, which is given as follows: suppose there exists a set of n different baseball cards, the collector wants to collect c complete sets of them, and the collector can obtain p unique cards in a "pack"at a time, with replacement. The first selection of a "pack"has time $t = 1$. The second selection has time $t = 2$ and so on. We want to find the probability of acquiring c complete sets after time t.

We define a state as an ordered tuple $(a_1, \ldots, a_c)$ which represents that a collector has collected exactly one copy of a_1 cards, exactly two copies of a_2 cards, and so on, with c copies or more of a_c cards. Clearly for each state we must have $a_1 + a_2 + \cdots + a_c \leq n$.

Given that constraint, we will order the states as follows: let $X_0 = (0, 0, ...0)$. Now, if $X_i = (a_1, \ldots a_c)$, define X_{i+1} as

$$\begin{cases} (a_1+1, a_2, a_3, \dots a_c) & : \sum_{i=1}^{c} a_i < n \\ (a_1, \dots, 0, a_{j+1}+1, \dots a_c) & : \sum_{i=1}^{c} a_i = n \end{cases} \tag{1}$$

where j is the lowest i such that a_i is nonzero. For example, the ordering of states for $n = 2, c = 3$ is

$$\begin{aligned} ((0,0,0),(1,0,0), \quad &(2,0,0),(0,1,0),(1,1,0), \\ &(0,2,0),(0,0,1),(1,0,1), \quad (0,1,1),(0,0,2)) \end{aligned}$$

We have $\binom{n+c}{c}$ states in all, so we index them in this order from 0 to $\binom{n+c}{c} - 1$. Using these states, we can create a transition matrix between states where the entries are the chance of moving from one state to another after collecting a new pack of cards. The formula for the probability of moving from state $(a_1, \dots, a_c)$ to $(b_1, \dots, b_c)$ is

$$\frac{\binom{n-\sum_{k=1}^{c} a_k}{y_1} \left(\prod_{k=1}^{c-1} \binom{a_k}{y_{k+1}} \right) \binom{a_c}{p-\sum_{k=1}^{c} y_k}}{\binom{n}{p}} \tag{2}$$

where $b_c = a_c + y_c$, $b_{c-1} = a_{c-1} - y_c + y_{c-1}$, etc. with $b_1 = a_1 - y_2 + y_1$. In other words, y_k denotes the number of cards of which a collector has $k-1$ copies and gains one more of that card from the collected pack. The first term in the numerator gives the number of ways to get the first copy from the set cards of which the collector previously had no copies. Each successive term gives the number of ways to get another copy from the set of cards of which the collector previously had k copies, and the last term gives the number of ways to get more copies of cards of which a collector already has c copies. Finally, the denominator gives the number of possible packs of cards, making the formula a probability.

Thus, we can create a transition matrix $A_{n,c,p}$ where each entry a_{ij} is given by the formula above, with $(a_1, \dots, a_c)$ being state i and $(b_1, \dots, b_c)$ being state j in our ordering. For example, $A_{3,2,2}$ is

	$(0,0)$	$(1,0)$	$(2,0)$	$(3,0)$	$(0,1)$	$(1,1)$	$(2,1)$	$(0,2)$	$(1,2)$	$(0,3)$
$(0,0)$	0	0	1	0	0	0	0	0	0	0
$(1,0)$	0	0	0	1/3	0	2/3	0	0	0	0
$(2,0)$	0	0	0	0	0	0	2/3	1/3	0	0
$(3,0)$	0	0	0	0	0	0	0	0	1	0
$(0,1)$	0	0	0	0	0	2/3	1/3	0	0	0
$(1,1)$	0	0	0	0	0	0	1/3	1/3	1/3	0
$(2,1)$	0	0	0	0	0	0	0	0	2/3	1/3
$(0,2)$	0	0	0	0	0	0	0	1/3	2/3	0
$(1,2)$	0	0	0	0	0	0	0	0	1/3	2/3
$(0,3)$	0	0	0	0	0	0	0	0	0	1

Note that for $\binom{n}{p} > 1$, $A_{n,c,p}$ transition matrices will have non-integer entries. To eliminate the complications of fractions, whenever we discuss $A_{n,c,p}$ we will refer to $\binom{n}{p} \cdot A_{n,c,p}$ to make its entries integers.

Proposition 1. *The transition matrix $A_{n,c,p}$ is upper triangular.*

Proof. Let $X_i = (a_1, \dots a_c)$ and $X_j = (b_1, \dots b_c)$ be two states in $A_{n,c,p}$, with $j < i$. Notice that as the indices of states increases by (1), for the greatest k such that a_k is nonzero, a_k never decreases; it either increases or stays the same. Compare X_i and X_j to find the largest k such that $a_k \neq b_k$. Since $j < i$, we must have $a_k > b_k$. Consider the transition from X_i to X_j. Transitioning between these states would then imply that the number of cards of which the collector has k copies decreases, but as $a_l = b_l$ for $l > k$, the number of cards of which the collector has more than k copies does not increase, a contradiction. Thus it is impossible for this transition to occur, making its probability 0.

So for any $j < i$, the entry in the ith row and jth column of $A_{n,c,p}$ is 0, implying that $A_{n,c,p}$ is upper triangular. □

1.2 Previous Results

The case of $c = 1$ has already been explored by Calkin and Edds. They considered the transition matrix $A_{n,1,p}$. To find the probability of having one complete set after time t they had to consider the probability of getting from state (0) to state (n) in $(A_{n,1,p})^t$. They used its Jordan Normal Form to compute this probability since $(A_{n,1,p})^t = P_{n,1,p} \cdot (J_{n,1,p})^t \cdot (P_{n,1,p})^{-1}$. They found that $A_{n,1,p}$ is an $n \times n$ matrix with n linearly independent eigenvectors making its Jordan Normal Form a diagonal matrix which simplified things. By computing $P_{n,1,p} \cdot (J_{n,1,p})^t \cdot (P_{n,1,p})^{-1}$ they were able to get the following expression for the probability of collecting one complete set of n cards taken p at a time after time t:

$$\sum_{i=0}^{n} \binom{n}{i} \left(\frac{\binom{n-i}{p}}{\binom{n}{p}} \right)^{t} (-1)^{i}$$

[?]

We will employ the same method to find the probability of getting two complete sets of n cards taken 1 at a time after time t.

2 Method to Find P

We first must to be able to find the Jordanizing Matrix P for a matrix A that satisfies $A = P \cdot J \cdot P^{-1}$ where J is the Jordan Normal Form of A.

Given a square matrix A, we can find its Jordan Normal Form J. For each eigenspace in J, on each block, we can find eigenvectors of A by solving the following system for v:

$$(A - \lambda I)v = 0.$$

We then find generalized eigenvectors up to the block size b by solving $(A - \lambda I)v_{i+1} = v_i$ with $1 \leq i \leq b - 1$ and $(A - \lambda I)v_1 = v$ where v is our original eigenvector corresponding to λ.

We can now construct P from these vectors (each vector being a column of P). These columns are given in terms of parameters.

3 $A_{n,2,1}$

We now consider a subset of the transition matrices $A_{n,c,p}$. We will be looking at the Jordanizing of $A_{n,2,1}$. First we must re-index our matrix making it easier to define the entries of the matrices in which we are interested.

3.1 Re-indexing our Matrices

Each $A_{n,2,1}$ has size $N = \frac{(n+1)(n+2)}{2}$ (i.e. a triangular number). Due to this fact and the behavior of our $n, 2, 1$ matrices, it is more helpful (for the purpose of defining the entries of the matrix) to index $A_{n,2,1}$,

$P_{n,2,1}$, $(P_{n,2,1})^{-1}$, and $J_{n,2,1}$ in the following way:

We first index the row position of the entry. Starting with $x = 0$ counting from the bottom, x will represent which horizonal block contains the entry. So $x = 0$ will be the bottom horizontal block, and $x = n$ is the top horizontal block. Each x consists of $x + 1$ rows. So we must have a second index that denotes the rows in each horizontal block x. Within each x, i will denote which row of x the entry is in. The bottom row of horizontal block x will be marked $i = 0$, and the top row of horizontal block x will be marked $i = x$.

Similarly, we index the column position of the entry. Starting with $y = 0$ counting from the right-hand side, y will denote which vertical block the entry is in. So $y = 0$ is the right-most vertical block, and $y = n$ is the left-most vertical block. Each y consists of $y+1$ columns. We must have a second index which denotes the columns in each vertical block y. Within each y, j will denote which column of y the entry is in. The right-most column of vertical block y is marked $j = 0$, and the left-most column of vertical block y will be marked $j = y$.

We now have 4 indices counting from the bottom right of our matrix rather than 2 indices counting from the top left. So each entry must be denoted $a_{x,i,y,j}$. For example take $n = 2$:

$$A_{2,2,1} = \left(\begin{array}{ccc|cc|c}
a_{2,2,2,2} & a_{2,2,2,1} & a_{2,2,2,0} & a_{2,2,1,1} & a_{2,2,1,0} & a_{2,2,0,0} \\
a_{2,1,2,2} & a_{2,1,2,1} & a_{2,1,2,0} & a_{2,1,1,1} & a_{2,1,1,0} & a_{2,1,0,0} \\
a_{2,0,2,2} & a_{2,0,2,1} & a_{2,0,2,0} & a_{2,0,1,1} & a_{2,0,1,0} & a_{2,0,0,0} \\
\hline
a_{1,1,2,2} & a_{1,1,2,1} & a_{1,1,2,0} & a_{1,1,1,1} & a_{1,1,1,0} & a_{1,1,0,0} \\
a_{1,0,2,2} & a_{1,0,2,1} & a_{1,0,2,0} & a_{1,0,1,1} & a_{1,0,1,0} & a_{1,0,0,0} \\
\hline
a_{0,0,2,2} & a_{0,0,2,1} & a_{0,0,2,0} & a_{0,0,1,1} & a_{0,0,1,0} & a_{0,0,0,0}
\end{array}\right)$$

Given an entry $a_{x,i,y,j}$ in $A_{n,2,1}$, we can find its traditional indexing (i.e. $a_{p,q}$, beginning our counting at 0):

$$p = \binom{n+2}{2} - 1 - i - \binom{x+1}{2}$$

$$q = \binom{n+2}{2} - 1 - j - \binom{y+1}{2}$$

By looking at the Jordan Normal Form $J_{n,2,1}$ of the integral $A_{n,2,1}$ this new indexing arises naturally.

Proposition 2. *Each eigenvalue λ is in a single Jordan block of size $n+1-\lambda$. So $J_{n,2,1}$ will look like:*

$$
\left(\begin{array}{ccccc|ccccc|cc|cc|c}
0 & 1 & & & & & & & & & & & & & \\
& \ddots & \ddots & & & & & & & & & & & & \\
& & \ddots & \ddots & & & & & & & & & & & \\
& & & \ddots & 1 & & & & & & & & & & \\
& & & & 0 & & & & & & & & & & \\
\hline
& & & & & 1 & 1 & & & & & & & & \\
& & & & & & \ddots & \ddots & & & & & & & \\
& & & & & & & \ddots & \ddots & & & & & & \\
& & & & & & & & \ddots & 1 & & & & & \\
& & & & & & & & & 1 & & & & & \\
\hline
& & & & & & & & & & \ddots & & & & \\
& & & & & & & & & & & \ddots & & & \\
\hline
& & & & & & & & & & & & n-1 & 1 & \\
& & & & & & & & & & & & & n-1 & \\
\hline
& & & & & & & & & & & & & & n
\end{array}\right)
$$

The proof of this proposition comes for free with the proof of Theorem 1 that appears later in the paper.

This behavior of the blocks of $J_{n,2,1}$ is motivation for this new system of "block indexing" because we can divide the matrix into horizontal blocks labelled x and vertical blocks labelled y, and when $x = y$, their intersection contains a Jordan block in $J_{n,2,1}$.

3.2 Probabilities for $A_{n,2,1}$

Using the indexing as discussed above, we can derive formulas for $A_{n,2,1}$, $J_{n,2,1}$, $P_{n,2,1}$, and $Q_{n,2,1} = (P_{n,2,1})^{-1}$. These formulas will allow us to find the probability of collecting two complete sets of n cards taken 1 at a time after time t. We can apply our formula for the entries of a transition matrix to find the values in $A_{n,2,1}$.

Lemma 1. *The entries of $A_{n,2,1}$ in terms of these new indices are given by:*

$$A_{n,2,1} = \left(a_{x,i,y,j}\right)$$

$$a_{x,i,y,j} = \binom{i}{i-j}\binom{x-i}{x-y}\binom{n-x}{1-(i-j)-(x-y)}$$

Proof. In $A_{n,2,1}$ each entry corresponds to the probability of moving from a state (a_1, a_2) to a state (b_1, b_2) multiplied by n.

For $A_{n,2,1}$ the probability of $(a_1, a_2) \to (b_1, b_2)$ is given by equation 2 as

$$\frac{\binom{n-(a_1+a_2)}{y_1}\binom{a_1}{y_2}\binom{a_2}{1-(y_1)-(y_2)}}{n}$$

Using the indices x,i,y,and j, the entry $a_{x,i,y,j}$ corresponds to the probability of moving from state $(x-i, n-x)$ to state $(y-j, n-y)$ multiplied by n. Therefore we get $y_1 = i-j$ and $y_2 = x-y$. So, the entries of $A_{n,2,1}$ are given by

$$a_{x,i,y,j} = \binom{n-(x-i+n-x)}{i-j}\binom{x-i}{x-y}\binom{n-x}{1-(i-j)-(x-y)}$$

□

Theorem 1. *The entries of $J_{n,2,1}$ in terms of these new indices are given by:*

$$J_{n,2,1} = \left(e_{x,i,y,j}\right)$$

$$e_{x,i,y,j} = \binom{x-y}{y-x}\binom{1}{i-j}\binom{n-x}{1-(i-j)}$$

The entries of the Jordanizing matrix $P_{n,2,1}$ in terms of these new indices are given by:

$$P_{n,2,1} = \left(p_{x,i,y,j}\right)$$

$$p_{x,i,y,j} = (j!)\binom{x-j}{y-j}\binom{i}{j}$$

The entries of $Q_{n,2,1} = (P_{n,2,1})^{-1}$ *in terms of these new indices are given by:*

$$Q_{n,2,1} = \left(q_{x,i,y,j}\right)$$

$$q_{x,i,y,j} = \frac{(-1)^{i+j+x+y}\binom{x-i}{x-y}\binom{i}{j}}{(i)!}$$

These formulas give us

$$A_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1} \cdot Q_{n,2,1}$$

Proof. It will suffice to show:

- $J_{n,2,1}$ is a block diagonal matrix made up of Jordan Blocks
- $A_{n,2,1} \cdot P_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1}$
- $P_{n,2,1} \cdot Q_{n,2,1} = I$

Our formula for each entry $e_{x,i,y,j}$ of $J_{n,2,1}$ is

$$\binom{x-y}{y-x}\binom{1}{i-j}\binom{n-x}{1-(i-j)}$$

The term $\binom{x-y}{y-x}$ in our formula is there to ensure that only when our horizontal and vertical block indices x and y (respectively) are equal do we have non-zero entries. Otherwise, if $x \neq y$, then either $x-y$ or $y-x$ is negative making $\binom{x-y}{y-x} = 0$.

The term $\binom{1}{i-j}$ is there to make it so only when $0 \leq i-j \leq 1$ do we have non-zero entries. Otherwise, $i-j < 0$ or $1 < i-j$ making $\binom{1}{i-j} = 0$.

The term $\binom{n-x}{1-(i-j)}$ is what gives $J_{n,2,1}$ the eigenvalues along the main diagonal. We know that $x = y$, and either $i-j = 0$ or $i-j = 1$ must hold in order to have non-zero entries. So only on the main and super diagonals of blocks where $x = y$, do we have non-zero entries. This makes $J_{n,2,1}$ a block diagonal matrix. When $x = y$ and $i-j = 1$, the entry is on the super diagonal of that block, and the formula gives us $e_{x,i,y,j} = 1$, following the description of a Jordan Normal Form.

So we have a block diagonal matrix with entries along the main diagonal of each block, and 1's along the super diagonal of each block making it a Jordan Normal Form.

Showing that $A_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1} \cdot Q_{n,2,1}$ and $(P_{n,2,1})^{-1} = Q_{n,2,1}$ will show that our formula for $J_{n,2,1}$ gives us the Jordan Normal Form of $A_{n,2,1}$ since the Jordan Normal Form of a matrix is unique up to block rearrangement.

$A_{n,2,1} \cdot P_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1}$ and $P_{n,2,1} \cdot Q_{n,2,1} = I$ is equivalent to $A_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1} \cdot Q_{n,2,1}$.

We will now show that $A_{n,2,1} \cdot P_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1}$.

Using the new indices, matrix multiplication will look like:

$$A_{n,2,1} \cdot P_{n,2,1} = \left(c_{x,i,y,j}\right)$$

$$c_{x,i,y,j} = \sum_{z=0}^{n} \sum_{k=0}^{z} a_{x,i,z,k} \cdot p_{z,k,y,j}$$

And

$$P_{n,2,1} \cdot J_{n,2,1} = \left(d_{x,i,y,j}\right)$$

$$d_{x,i,y,j} = \sum_{z=0}^{n} \sum_{k=0}^{z} p_{x,i,z,k} \cdot e_{z,k,y,j}$$

Now we can use our formulas for our matrices to see what these double-sums look like.

$$c_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z}\binom{i}{i-k}\binom{x-i}{x-z}\binom{n-x}{1-(i-k)-(x-z)}(j!)\binom{z-j}{y-j}\binom{k}{j}$$

$$d_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z}(k!)\binom{x-k}{z-k}\binom{i}{k}\binom{z-y}{y-z}\binom{1}{k-j}\binom{n-z}{1-(k-j)}$$

The way $\binom{\alpha}{\beta}$ is defined, if $\alpha < 0, \beta < 0$, or $\alpha < \beta$, $\binom{\alpha}{\beta} = 0$. Using this fact, we can find some constraints on z and k in both these double sums.

Take the sum representing $c_{x,i,y,j}$, the entry $\binom{i}{i-k}$ shows that $k \leq i$ otherwise $i-k < 0$ making $\binom{i}{i-k} = 0$ therefore, any term with $k > i$ does not affect the double-sum. Now look at the entry $\binom{x-i}{x-z}$. This makes $z \leq x$ since if it were otherwise, $x - z < 0$ making $\binom{x-i}{x-z} = 0$. So any term with $z > x$ does not affect the double-sum.

With these two constraints in mind, consider the entry $\binom{n-x}{1-(i-k)-(x-z)}$. We know $k \leq i$ and $z \leq x$, therefore $i - k \geq 0$ and $x - z \geq 0$. So the only possible non-zero values of $\binom{n-x}{1-(i-k)-(x-z)}$ can be when $1-(i-k)-(x-z) = 0$ or 1 due to our constraints on z and k. So either both $(i-k)$ and $(x-z)$ are 0, or only one of them equals 1 and the other is 0. So there are only three pairs of values for z and k that give a non-zero term in our double-sum. The pairs are: $z = x$, $k = i-1$; $z = x$, $k = i$; and $z = x-1$, $k = i$.

So we have:

$$c_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z}\binom{i}{i-k}\binom{x-i}{x-z}\binom{n-x}{1-(i-k)-(x-z)}(j!)\binom{z-j}{y-j}\binom{k}{j}$$

$$= \binom{i}{1}(j!)\binom{x-j}{y-j}\binom{i-1}{j} + \binom{n-x}{1}(j!)\binom{x-j}{y-j}\binom{i}{j} + \binom{x-i}{1}(j!)\binom{x-1-j}{y-j}\binom{i}{j}$$

$$= (i)(j!)\binom{x-j}{y-j}\binom{i-1}{j} + (n-x)(j!)\binom{x-j}{y-j}\binom{i}{j} + (x-i)(j!)\binom{x-1-j}{y-j}\binom{i}{j}$$

We now focus on finding constraints for:

$$d_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z}(k!)\binom{x-k}{z-k}\binom{i}{k}\binom{z-y}{y-z}\binom{1}{k-j}\binom{n-z}{1-(k-j)}$$

Consider the entry $\binom{z-y}{y-z}$. If $z \neq y$, $\binom{z-y}{y-z} = 0$. So the only non-zero terms in this double-sum occur when $y = z$. Moreover, consider the entry $\binom{1}{k-j}$. In order for the term to be non-zero, $k - j$ must either be 1 or 0. So $k = j$ or $k = j + 1$.

So we have:

$$d_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z}(k!)\binom{x-k}{z-k}\binom{i}{k}\binom{z-y}{y-z}\binom{1}{k-j}\binom{n-z}{1-(k-j)}$$

$$= (j!)\binom{x-j}{z-k}\binom{i}{j}\binom{0}{0}\binom{1}{0}\binom{n-y}{1} + (j+1)!\binom{x-j-1}{y-j-1}\binom{i}{j+1}\binom{0}{0}\binom{1}{1}\binom{n-z}{0}$$

$$= (j!)\binom{x-j}{y-j}\binom{i}{j}(n-y) + (j+1)!\binom{x-j-1}{y-j-1}\binom{i}{j+1}$$

We can now show that these new constrained expressions are equal. First we simplify $c_{x,i,y,j}$.

$$c_{x,i,y,j} = (i)(j!)\binom{x-j}{y-j}\binom{i-1}{j} + (n-x)(j!)\binom{x-j}{y-j}\binom{i}{j}$$

$$+(x-i)(j!)\binom{x-1-j}{y-j}\binom{i}{j}$$

$$= \frac{(i)(j!)(x-j)!(i-1)!}{(y-j)!(x-y)!(j!)(i-j-1)!} + \frac{(n-x)(j!)(x-j)!(i!)}{(y-j)!(x-y)!(j!)(i-j)!}$$

$$+ \frac{(x-i)(j!)(x-j-1)!(i!)}{(y-j)!(x-y-1)!(j!)(i-j)!}$$

$$= \frac{(x-j)!(i!)}{(y-j)!(x-y)!(i-j-1)!} + \frac{(n-x)(x-j)!(i!)}{(y-j)!(x-y)!(i-j)!} + \frac{(x-i)(x-j-1)!(i!)}{(y-j)!(x-y-1)!(i-j)!}$$

$$= \frac{(x-j)!(i!)(i-j)}{(y-j)!(x-y)!(i-j)!} + \frac{(n-x)(x-j)!(i!)}{(y-j)!(x-y)!(i-j)!} + \frac{(x-i)(x-j-1)!(i!)(x-y)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(i!)(x-j-1)!\left((x-j)(i-j)+(n-x)(x-j)+(x-i)(x-y)\right)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(i!)(x-j-1)!(xi-xj-ji+j^2+nx-nj-x^2+xj+x^2-xy-xi+yi)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(i!)(x-j-1)!(-ji+j^2+nx-nj-xy+yi)}{(y-j)!(x-y)!(i-j)!}$$

Then we simplify $d_{x,i,y,j}$.

$$d_{x,i,y,j} = (j!)\binom{x-j}{y-j}\binom{i}{j}(n-y) + (j+1)!\binom{x-j-1}{y-j-1}\binom{i}{j+1}$$

$$= \frac{(j!)(x-j)!(i!)(n-y)}{(y-j)!(x-y)!(j!)(i-j)!} + \frac{(j+1)!(x-j-1)!(i!)}{(y-j-1)!(x-y)!(j+1)!(i-j-1)!}$$

$$= \frac{(x-j)!(i!)(n-y)}{(y-j)!(x-y)!(i-j)!} + \frac{(x-j-1)!(i!)(y-j)(i-j)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(x-j-1)!(i!)\left((x-j)(n-y)+(y-j)(i-j)\right)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(x-j-1)!(i!)(xn-xy-nj+jy+yi-yj-ij+j^2)}{(y-j)!(x-y)!(i-j)!}$$

$$= \frac{(i!)(x-j-1)!(-ji+j^2+nx-nj-xy+yi)}{(y-j)!(x-y)!(i-j)!}$$

So $c_{x,i,y,j} = d_{x,i,y,j}$. Implying $A_{n,2,1}P_{n,2,1} = P_{n,2,1}J_{n,2,1}$.

Now we must show $P_{n,2,1} \cdot Q_{n,2,1} = I$.

Let $P_{n,2,1} \cdot Q_{n,2,1} = (r_{x,i,y,j})$. Following the matrix multiplication for our system of indexing, we get

$$r_{x,i,y,j} = \sum_{z=0}^{n}\sum_{k=0}^{z} \frac{(k)!\binom{x-k}{z-k}\binom{i}{k}(-1)^{k+j+z+y}\binom{z-k}{z-y}\binom{k}{j}}{(k)!}$$

$$= \sum_{z=0}^{n}\sum_{k=0}^{z} \binom{x-k}{z-k}\binom{i}{k}(-1)^{k+j+z+y}\binom{z-k}{z-y}\binom{k}{j}$$

Notice that $P_{n,2,1}$ and $Q_{n,2,1}$ are upper triangular making $P_{n,2,1} \cdot Q_{n,2,1}$ upper triangular. So whenever $x < y$, $r_{x,i,y,j} = 0$. Also, consider the terms $\binom{i}{k}$ and $\binom{k}{j}$ in the double sum. These imply that $k \leq i$ and $j \leq k$ respectively. Otherwise, $r_{x,i,y,j} = 0$. Putting the two inequalities together we get that $j \leq i$ must hold, otherwise we would get $r_{x,i,y,j} = 0$. From this deduction, we know that $r_{x,i,y,j} = 0$ when $x < y$ or $i < j$. So we only need to consider cases where $y \leq x$ and $j \leq i$.

It will suffice to show that $r_{x,i,y,j} = 1$ when $x = y \wedge i = j$ (an entry on the main diagonal) and $r_{x,i,y,j} = 0$ when $x \neq y \vee i \neq j$ (an entry not on the main diagonal).

Let $x = y \wedge i = j$. By looking at the binomial coefficients just as before we can see that $y \leq z \leq x$ and $j \leq k \leq i$. Since $x = y$ we have $y = z = x$, and similarly we have $i = j = k$. We will now go through the expression for $r_{x,i,y,j}$ and keep every x while replacing y

and z with x, and we will also keep every i while replacing j and k with i. Doing this we get

$$r_{x,i,y,j} = \binom{x-i}{x-i}(-1)^{i+i+x+x}\binom{x-i}{x-x}\binom{i}{i}$$

$$= (1)(-1)^{2(i+x)}(1)(1)$$

$$= 1$$

So whenever $x = y \wedge i = j$ we get $r_{x,i,y,j} = 1$.

Next we show that if $x \neq y \vee i \neq j$ then $r_{x,i,y,j} = 0$. We will separate this into three cases.

We will be taking advantage of the two following identities:

$$\sum_{c=0}^{d}(-1)^c\binom{d}{c} = 0$$

$$\binom{\alpha}{\beta}\binom{\beta}{\gamma} = \binom{\alpha}{\gamma}\binom{\alpha-\gamma}{\beta-\gamma}$$

[**?**]

Case I: If $x = y \wedge i \neq j$ **then** $r_{x,i,y,j} = 0$

So as before we have $x = y = z$ so we have no need for the sum over z. We will just keep each x and replace each y and z with x, but the rest of the expression still has i, k, and j in it. Maintaining that $j \leq k \leq i$ we get

$$r_{x,i,y,j} = \sum_{k=j}^{i}\binom{x-k}{x-k}\binom{i}{k}(-1)^{k+j+x+x}\binom{x-k}{x-x}\binom{k}{j}$$

$$= \sum_{k=j}^{i}(-1)^{k+j+2x}\binom{i}{k}\binom{k}{j}$$

$$= \sum_{k=j}^{i}(-1)^{k+j}\binom{i}{j}\binom{i-j}{k-j}$$

Note that $(-1)^{k+j} = (-1)^{k-j}$, so we get

$$= \binom{i}{j} \sum_{k=j}^{i} (-1)^{k-j} \binom{i-j}{k-j}$$

We now let $s = i - j$ and $l = k - j$ and get

$$r_{x,i,y,j} = \binom{i}{j} \sum_{l=0}^{s} (-1)^{l} \binom{s}{l}$$

$$= 0$$

Which confirms Case I.

Case II: If $x \neq y \wedge i = j$ then $r_{x,i,y,j} = 0$

Now we have $i = j = k$ so we have no need for the sum over k. We will just keep each i and replace each j and k with i, but the rest of the expression will still have x, y, and z in it. So maintaining that $y \leq z \leq x$ we get

$$r_{x,i,y,j} = \sum_{z=y}^{x} \binom{x-i}{z-i} \binom{i}{i} (-1)^{i+i+z+y} \binom{z-i}{z-y} \binom{i}{i}$$

$$= \sum_{z=y}^{x} (-1)^{2i+z+y} \binom{x-i}{z-i} \binom{z-i}{z-y}$$

$$= \sum_{z=y}^{x} (-1)^{z+y} \binom{x-i}{z-y} \binom{(x-i)-(z-y)}{(z-i)-(z-y)}$$

Note that $(-1)^{z+y} = (-1)^{z-y}$, so we get

$$= \sum_{z=y}^{x} (-1)^{z-y} \binom{x-i}{z-y} \binom{x-i-(z-y)}{y-i}$$

We now let $z - y = w$ to give us

$$= \sum_{w=0}^{x-y} (-1)^{w} \binom{x-i}{w} \binom{x-i-w}{y-i}$$

$$= \sum_{w=0}^{x-y} (-1)^{w} \binom{x-i}{x-i-w} \binom{x-i-w}{y-i}$$

$$= \sum_{w=0}^{x-y} (-1)^w \binom{x-i}{y-i} \binom{x-i-(y-i)}{x-i-w-(y-i)}$$

$$= \binom{x-i}{y-i} \sum_{w=0}^{x-y} (-1)^w \binom{x-y}{x-w-y}$$

$$= \binom{x-i}{y-i} \sum_{w=0}^{x-y} (-1)^w \binom{x-y}{w}$$

We now let $x - y = v$ so we get

$$r_{x,i,y,j} = \binom{x-i}{y-i} \sum_{w=0}^{v} (-1)^w \binom{v}{w}$$

$$= 0$$

Which confirms Case II.

Case III: If $x \neq y \wedge i \neq j$ then $r_{x,i,y,j} = 0$

We know that $y \leq z \leq x$ and $j \leq k \leq i$, so we have that

$$r_{x,i,y,j} = \sum_{z=y}^{x} \sum_{k=j}^{i} (-1)^{k+j+z+y} \binom{x-k}{z-k} \binom{i}{k} \binom{z-k}{z-y} \binom{k}{j}$$

Note that $(-1)^{k+j+z+y} = (-1)^{k-j}(-1)^{z-y}$ giving us

$$= \sum_{z=y}^{x} \sum_{k=j}^{i} (-1)^{k-j}(-1)^{z-y} \binom{i}{k} \binom{k}{j} \binom{x-k}{z-k} \binom{z-k}{z-y}$$

$$= \sum_{z=y}^{x} \sum_{k=j}^{i} (-1)^{k-j}(-1)^{z-y} \binom{i}{j} \binom{i-j}{k-j} \binom{x-k}{z-k} \binom{z-k}{z-y}$$

Let $l = k - j$ and $s = i - j$ giving us

$$= \binom{i}{j} \sum_{z=y}^{x} \sum_{l=0}^{s} (-1)^{z-y}(-1)^{l} \binom{s}{l} \binom{x-l-j}{z-l-j} \binom{z-l-j}{z-y}$$

Now let $w = z - y$

$$= \binom{i}{j} \sum_{w=0}^{x-y} (-1)^w \sum_{l=0}^{s} (-1)^l \binom{s}{l} \binom{x-l-j}{w+y-l-j} \binom{w+y-l-j}{w}$$

Let $v = x - y$

$$= \binom{i}{j} \sum_{w=0}^{v} (-1)^w \sum_{l=0}^{s} (-1)^l \binom{s}{l} \binom{v+y-l-j}{w+y-l-j} \binom{w+y-l-j}{y-l-j}$$

$$= \binom{i}{j} \sum_{w=0}^{v} (-1)^w \sum_{l=0}^{s} (-1)^l \binom{s}{l} \binom{v+y-l-j}{y-l-j} \binom{v+(y-l-j)-(y-l-j)}{w+(y-l-j)-(y-l-j)}$$

$$= \binom{i}{j} \sum_{w=0}^{v} (-1)^w \sum_{l=0}^{s} (-1)^l \binom{s}{l} \binom{v+y-l-j}{v} \binom{v}{w}$$

$$= \binom{i}{j} \left(\sum_{w=0}^{v} (-1)^w \binom{v}{w} \right) \left(\sum_{l=0}^{s} (-1)^l \binom{s}{l} \binom{v+y-l-j}{v} \right)$$

$$= 0$$

Confirming Case III. Recall that if $x < y \vee i < j$ then we have $r_{x,i,y,j} = 0$. So we have that whenever $x \neq y \vee i \neq i$ is true, $r_{x,i,y,j} = 0$.

So when $x = y \wedge i = j$ we have $r_{x,i,y,j} = 1$, and when $x \neq y \vee i \neq j$ we have $r_{x,i,y,j} = 0$. Recall that we let $(r_{x,i,y,j}) = P_{n,2,1} \cdot Q_{n,2,1}$. So this shows that $P_{n,2,1} \cdot Q_{n,2,1} = I$.

We have shown:

- $P_{n,2,1} \cdot Q_{n,2,1} = I$
- $A_{n,2,1} \cdot P_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1}$
- Our formula for $J_{n,2,1}$ creates the Jordan Normal Form of $A_{n,2,1}$ since it makes a matrix that satisfies the requirements of being a Jordan Normal Form and every matrix has a unique Jordan Normal Form up to block rearrangement.

Which tells us that our formulas are correct for these matrices such that $A_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1} \cdot Q_{n,2,1}$. $\square$

3.3 Computing $(A_{n,2,1})^t$

Now that we have a definite, closed formula for both $P_{n,2,1}$ and $(P_{n,2,1})^{-1}$ in the formula $A_{n,2,1} = P_{n,2,1} \cdot J_{n,2,1} \cdot (P_{n,2,1})^{-1}$, we can look at the probability of collecting c complete sets after time t . Since $A_{n,2,1}$ is just the probability of transitioning from one state to another, we need to look at the matrix $(A_{n,2,1})^t = P_{n,2,1} \cdot (J_{n,2,1})^t \cdot (P_{n,2,1})^{-1}$ to find our probability. The desired entry will occur in the upper right corner, as we want to transition from the first state to the last state in t tries. However, we still need a closed formula for $(J_{n,2,1})^t$ in terms of our x, i, y, j indexing.

Proposition 3. *If* $(J_{n,2,1})_{x,i,y,j} = \binom{x-y}{y-x}\binom{1}{i-j}\binom{n-x}{1-(i-j)}$, *then we have*

$$((J_{n,2,1})^t)_{x,i,y,j} = \binom{x-y}{y-x}\binom{t}{i-j}(n-x)^{t-(i-j)}$$

Proof. It's clear that we can consider each Jordan block seperately in our exponentiation of $J_{n,2,1}$. Thus the term $\binom{x-y}{y-x}$ remains to ensure only blocks on the main diagonal contain non-zero entries. Next, a well known formula for taking a Jordan block to a power basically gives the terms of $(\lambda+1)^t$, starting on each entry of the main diagonal with the highest power of λ and going right, stopping at the end of the Jordan block. This is of course the terms of the binomial expansion of degree t, and the power of each must be $i-j$ by our indexing. Finally, the eigenvalue is given as $n-x$. Thus we have the above formula. $\square$

With this formula, we can now find a better formula for the upper-right corner entry of $(A_{n,2,1})^t$, using $P_{n,2,1} \cdot (J_{n,2,1})^t \cdot (P_{n,2,1})^{-1}$.

Theorem 2. *The probability of collecting* 2 *complete sets of* n *cards taken* 1 *at a time after time* t *is*

$$\sum_{z=0}^{n}(-1)^z \sum_{l=0}^{z}\left(\frac{\frac{n!}{(z-l)!(n-z)!}\binom{t}{l}(n-z)^{t-l}}{n^t}\right)$$

Proof. We will associate $P_{n,2,1}$ and $(J_{n,2,1})^t$ first, then multiply the result by $(P_{n,2,1})^{-1}$. Further we will limit our look to only the upper right entry of $P_{n,2,1} \cdot (J_{n,2,1})^t \cdot (P_{n,2,1})^{-1}$, so we need only take the top row of $P_{n,2,1}$ times every column of $(J_{n,2,1})^t$, and then the top row of $P_{n,2,1} \cdot (J_{n,2,1})^t$ times the right-most column of $(P_{n,2,1})^{-1}$. By our x, i, y, j indexing this is

$$\sum_{z=0}^{n}\sum_{k=0}^{z}\left[\sum_{w=0}^{n}\left(\sum_{l=0}^{w}(P_{n,2,1})_{n,n,w,l}((J_{n,2,1})^t)_{w,l,z,k}\right)((P_{n,2,1})^{-1})_{z,k,0,0}\right]$$

However, if $w \neq z$, $((J_{n,2,1})^t)_{w,l,z,k} = 0$, so this reduces to

$$\sum_{z=0}^{n}\sum_{k=0}^{z}\left[\sum_{l=0}^{z}\left((P_{n,2,1})_{n,n,z,l}\cdot((J_{n,2,1})^t)_{z,l,z,k}\right)((P_{n,2,1})^{-1})_{z,k,0,0}\right]$$

We now plug in all of our formulas to obtain

$$\sum_{z=0}^{n}\sum_{k=0}^{z}\left[\sum_{l=0}^{z}\left((l!)\binom{n-l}{z-l}\binom{n}{l}\binom{t}{l-k}(n-z)^{t-(l-k)}\right)\frac{(-1)^{z+k}\binom{z-k}{z}}{k!}\right].$$

We know that $k \geq 0$, so we will have $\binom{z-k}{z} = 0$ if $k > 0$. So $k = 0$ is the only value of k that gives a non-zero term in the outer sum. So a final reduction gives

$$\sum_{z=0}^{n}(-1)^z\sum_{l=0}^{z}\left(\frac{n!}{(z-l)!(n-z)!}\binom{t}{l}(n-z)^{t-l}\right)$$

Finally, to create $A_{n,c,p}$, we multiplied the original transition matrix by $\binom{n}{p}$, so we need to divide our final result by n^t to get the final probability.

□

4 Further Work and Conjectures

Following our work from the $n,2,1$ case, we attempted to find similar relations and patterns for higher c. However, once c is greater than 2, the eigenspaces begin splitting up and become multidimensional. Our program can solve for the P with multidimensional eigenspaces, but the patterns within have proven too difficult to recreate. However, we have conjectures about the Jordan Normal Form for $n,3,1$ and $n,4,1$. This preliminary conjecture will help set up those ideas.

Conjecture 1. *For each eigenspace in $J_{n,c,1}$, there exists an eigenspace of the same dimension with the same block sizes, with an eigenvalue one above the original, in $J_{n+1,c,1}$.*

So for $n, c, 1$, increasing n by 1 will send all of the eigenspaces to the same eigenspace in $n+1, c, 1$, with the eigenvalue increased by 1. Naturally there will also be an additional eigenspace added for the 0 eigenvalue. As the eigenspaces can then be defined recursively, the following conjectures will only describe the new eigenspace, for eigenvalue 0, given a certain n.

Conjecture 2. *For $J_{n,3,1}$, the eigenspace over eigenvalue 0 contains blocks of size $2n+1, 2n-3, 2n-7, \ldots$.*

Conjecture 3. *For $J_{0,4,1}$, the eigenspace over eigenvalue 0 contains 1 block of size 1. For $J_{n,4,1}, n > 0$, the eigenspace over eigenvalue 0 contains all blocks from the 0 eigenspace of $J_{n-1,4,1}$, except each block is 3 sizes larger, along with blocks of size $n+1, n-3, n-7, \ldots$, with the exception that a block of size 2 will never be created.*

The following table shows some block sizes for the 0 eigenspace over certain n and c, computionally determined:

	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
$c=1$	1	1	1	1	1	1
$c=2$	1	2	3	4	5	6
$c=3$	1	3	5, 1	7, 3	9, 5, 1	11, 7, 3
$c=4$	1	4	7, 3	10, 6, 4	13, 9, 7, 5, 1	16, 12, 10, 8, 6, 4
$c=5$	1	5	9, 5, 1	13, 9, 7, 5, 1	17, 13, 11, 9, 9, 5, 5, 1	21, 17, 15, 13, 13, 11, 9, 9, 7, 5, 5, 1

Conjecture 4. *The block sizes for the 0 eigenspace of $J_{n,c,1}$ are the same as the block sizes for the 0 eigenspace of $J_{c-1,n+1,1}$.*

Acknowledgements

- Kevin James, Jeff Beyerl, Matt Saltzman, Jim Brown, Jonathan Hanke, Dave Penniston, Chantal David, Benoit Larose, Mohammed Tesemma, Robert Schneider, Yang Wang
- Clemson University
- National Science Foundation