Final Project: College Admissions Early/Regular Decision Analysis

Walter Shen

APM 115: Mathematical Modeling

Abstract

The college application process in the United States is seen as a ritual of adulthood for many people. Students apply to various colleges; they strategize the application to maximize success. Colleges also have to optimize student acceptance and enrollment in a similar fashion. One method that is employed is the utilization of Early Action (EA) and Early Decision (ED) programs, in conjunction to the traditional Regular Decision (RD) program at colleges in the United States.

In our analysis, we will analyze and discover the various effects of different factors in ED and RD admissions.

Section 1: Introduction and Importance

Early admissions programs are widely employed by colleges to get more prospective students to become more enthusiastic about a specific college[1]. However, given the various college choices a student has, students often apply to several colleges, in hopes of maximizing success. A side effect of this newer phenomenon is on the shoulders of the college admissions officers, who must be able to maintain the student population size without compromising quality of the students. The student yield—the proportion of admitted students who actually attend an institution—can be rather low for certain colleges, making some college admissions offices anxious. One way colleges artificially inflate yield is through the usage of Early Action (EA) admissions, where students can apply to a college (and receive a decision) months before normal application deadlines in the spring. Early Action applications often signify genuine interest; however, Early Action decisions are non-binding, as are Regular Decision (RD) decisions, containing the same yield phenomenon. Early Decision (ED) programs, however, are binding, meaning that once a student is accepted, he or she must attend that college.

Many colleges use Early Decision to solidify more of their student body earlier on. At many schools, applying ED often results in a higher rate of acceptance [2]—meaning, a student who applied ED to a school might not have been admitted to the school had he or she applied in the RD round. This results in fears of a lower quality student body if a college accepts at a higher rate through the ED round. Further, over-enrollment[3] is a problem colleges fear to face, as is under-enrollment[4]. We will be analyzing these features in our model.

Section 2: Description of the Model

2.1 Single College Model

To attempt to model this college admissions phenomenon, we can create a Markov chain to represent the flow of students between colleges' application processes, at different points in time.

We can have each state represent applicants at different stages in the college application process (e.g. initial pool, attending). In a single college we can create a following Markov chains with different states:

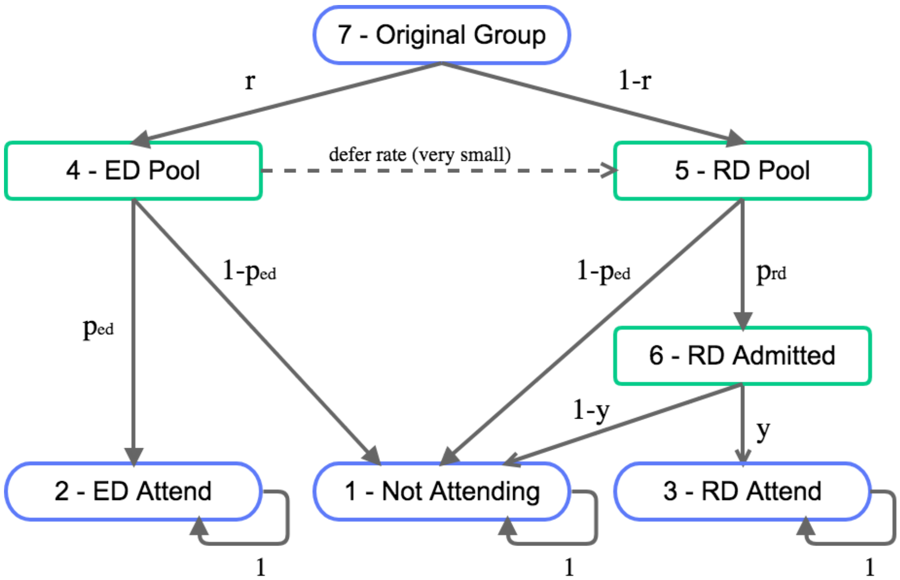


Figure 1: Single College Markov Chain with numbered states

Visualized are the numbered states. Students in the Original Group either apply ED or RD with a certain proportion, and the people in those pools get admitted according to their respective proportions. ED is binding, so all students that get admitted will attend. RD is nonbinding, so students will eventually attend with a yield rate. Students who get rejected ED or RD, or choose not to attend after getting admitted RD, will get absorbed by a Not Attending state. There is a very small (and for the purposes of this project, negligible) defer rate between the ED pool and the RD pool.

As such, we can define a transition matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 1-p_{ed} & 1-p_{rd} & 1-y & 0 \\ 0 & 1 & 0 & p_{ed} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & r \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-r \\ 0 & 0 & 0 & 0 & p_{rd} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A primary motivation for ED applicants is a higher rate of acceptance than in the RD pool. As such, we can modify our transition matrix by allowing the proportion the proportion r at which the Original Group flows into the ED Pool to be proportional to the difference in p_{ed} and prd:

$$r \propto p_{ed} - p_{rd}$$

This proportion is only if $p_{ed} - p_{rd}$ is positive.

This method allows us to have the final proportion of states be limited to just three variables: the ED acceptance rate, the RD acceptance rate, and the yield rate.

2.2 Two College Model

In another simplified model, we can compare two selective colleges in close geographic proximity. These two selective colleges are directly competing over a student pool. To be able to make this simplified model, we make two assumptions:

1) All students in this pool (eg. high school, county) will apply to one college ED.
    a) We can also imply that this pool is a competitive pool; and as such, will want to maximize their success with applying ED and then RD if needed.
2) They will apply RD only if rejected in the ED round.

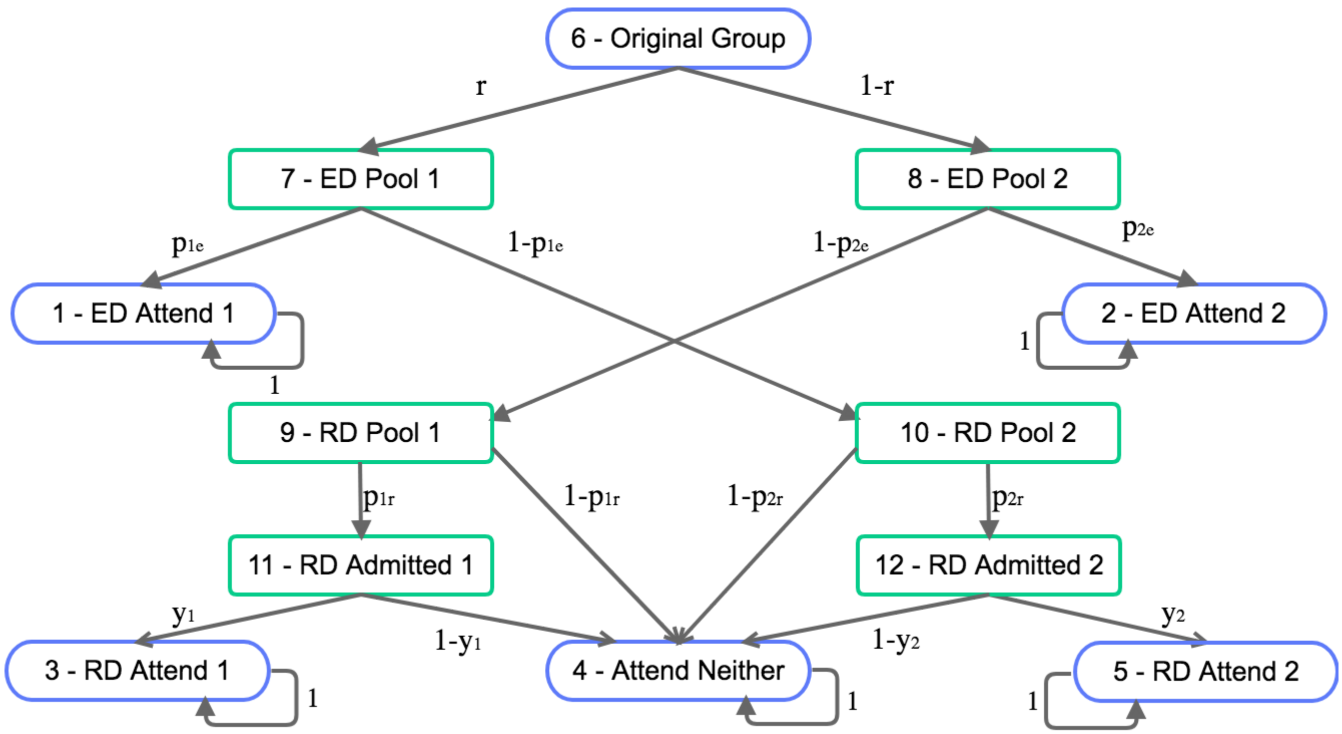


Figure 2: Two College Markov Chain with numbered states

The original student pool in the starting state will choose to apply ED to either College 1 or College 2, and depending on the result of the ED round, they will either attend College 1 or College 2, or apply to the college RD they did not apply to in ED. Following this, they will get accepted (and some will attend at a yield rate) or will not attend either college.

We can define a transition matrix with these states:

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & p_{1e} & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & p_{2e} & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1-p1r & 1-p_{2r} & 1-y_1 & 1-y_2 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & y_2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & r & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1-r & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p_{2e} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1-p_{1e} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{1r} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{2r} & 0 & 0
\end{bmatrix}$$

We can compare the proportion of students choosing to apply to College 1 over College 2 as proportional to their ED admit rates relative to the other college:

$$r \propto \frac{p_{1e}}{p_{1e}+p_{2e}}$$

This proportion is only valid if $\frac{p_{1e}}{p_{1e}+p_{2e}}$ is positive.

This method allows us to have the final proportion of states be limited to just the ED acceptance rate, the RD acceptance rate, and the yield rate of both colleges.

Section 3: Analysis and Results

3.1 Single College Model

Our first model in Figure 1 is very literally interpreted by the American college admissions process, so we can quickly use this Markov Chain to demonstrate a university's process:

We look at freshman undergraduate admission data from Cornell University, Ithaca, NY (information gathered from Cornell's 2017-18 Common Data Set (CDS)[5]):

	Cornell	Model
Prop in state 1 (not attend)	0.9288	0.9304
Prop in state 2 (ED attend)	0.0296	0.0275
Prop in state 3 (RD attend)	0.0416	0.0421

Table 1: Model results using Cornell CDS data

Note that errors encountered in this model are partially due to rounding errors, as well as exceptions in the ED-RD system at Cornell (and at other colleges in the US), such as a difference in the ED round (for reconsideration in RD), other students break the binding agreement of the ED round for financial reasons, rescindment of acceptances (and the student not attending the college), or other factors.

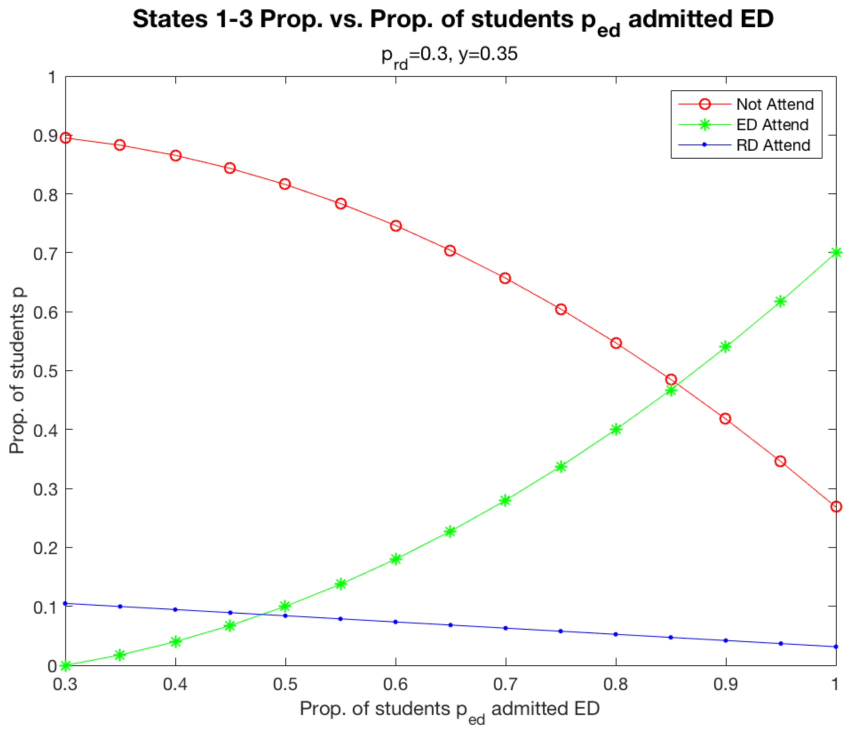
As such, values for proportions for Cornell University are estimated by assuming no students were deferred from ED to RD, and that all students admitted ED were admitted RD. The p_{ed}, p_{rd}, r, and yield values were determined from this method, and were plugged in the Markov Chain Model to produce the proportions of the states.

We can now model an example where the proportion r where students choose to apply ED is proportional to the advantage advantage of ED applicants (we let the coefficient of proportionality equal to 1). Allowing p_{ed} = 0.4, p_{rd} = 0.3, and y = 0.35:

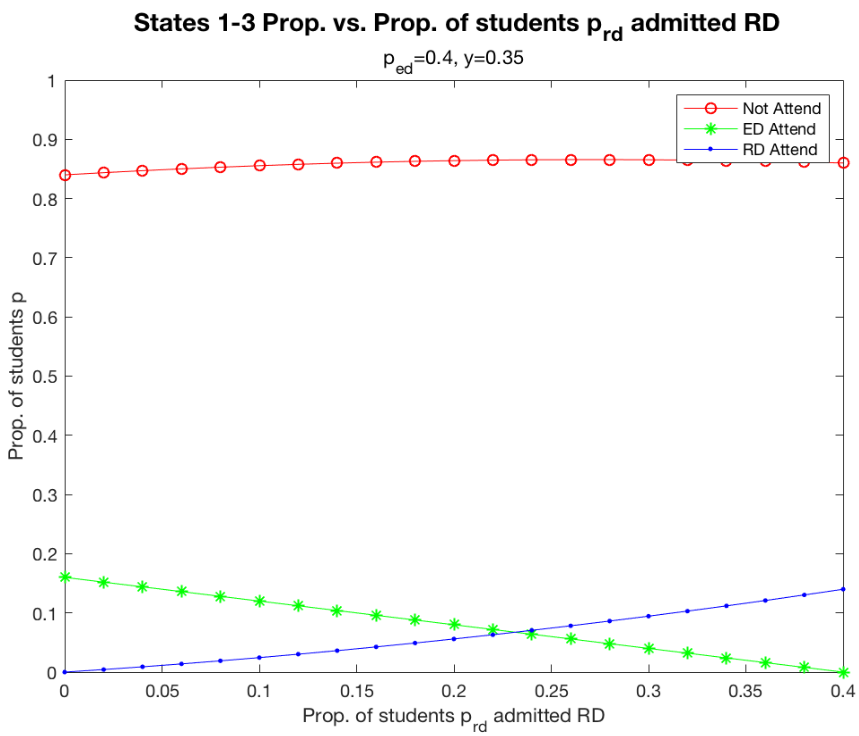
Proportion of original group not attending:	0.8655
Proportion of original group attending via ED:	0.04
Proportion of original group attending via RD:	0.0945
Proportion of attending getting in via ED:	0.2974

Table 2: Results of modified (r proportional to advantage), with p_{ed} = 0.4, p_{rd} = 0.3, and y = 0.35

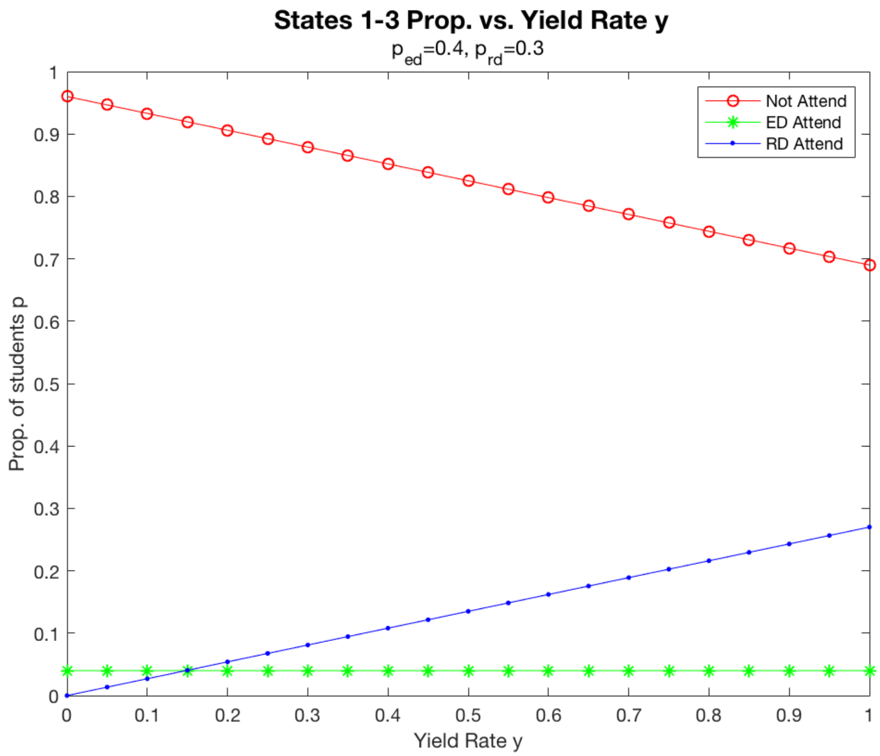
We can model the effect of p_{ed}, p_{rd}, and y on the final proportions by varying these values (while keeping the other two constant) and graphing:



As the proportion of students p_{ed} admitted in the ED round increases, then the proportion of the original student population attending via the ED round increases in a non-linear path (partially due to the proportion r applying via ED being proportional to the advantage of applying ED); the non-attending group decreases due to binding applications; RD attendance lowers because of most students attending being from the ED round.



As the proportion of students p_{ed} admitted in the RD round increases (catching up to the ED rate, after which is irrelevant to our study, as RD rates inherently are higher than ED rates), then the proportion of the original student population attending via the RD round increases in a non-linear path. If more students are accepted, more will decide to attend from the RD admittance.



As the yield rate y increases more students admitted in the RD round will attend the college at a linear rate; this is due to the effects of a non-binding agreement (RD) being lowered with a higher yield rate. Therefore, the proportion of students attending from the RD round easily surpasses the ED rate, especially with a size boost from the start.

The analysis of the yield rate y lends itself to further analysis. Yield is a factor that college admissions officers often anxiously anticipate, due to possibility of over- or under-enrollment[6].

Due to a lack of analyzable data that may be easily and accurately used for a college's yield over time, we make a (risky but needed) assumption that the yield rate follows a normal distribution, with mean y_0 and standard deviation σ_y.

The total enrollment of a college can be calculated from the Markov Chain model as a function of y:

$$Enrollment(y) = r \cdot p_{ed} + (1 - r) \cdot p_{rd} \cdot y$$

We can calculate the mean and variance of the enrollment function:

$$Mean(Enr(y)) = r \cdot p_{ed} + (1 - r) \cdot p_{rd} \cdot y_0$$
$$Var(Enr(y)) = (1 - r)^2 \cdot {p_{rd}}^2 \cdot {\sigma_y}^2$$
$$Sd(Enr(y)) = (1 - r) \cdot p_{rd} \cdot \sigma_y$$

To optimize this situation we wish to minimize $p_{ed} - p_{rd}$ (to maintain student quality consistency between ED and RD attending students, and have each have similar caliber by having similar selectivity) subject to: 1) keeping $Sd(Enr(y))$—standard deviation—below a threshold, and 1) keeping $Mean(Enr(y))$, average enrollment equal to a constant.

Using the `linprog` linear programming solver in MATLAB we are be able to optimize this situation with one example:

Proportion applying ED:	0.3
Standard deviation of yield rate:	0.03
Average yield rate:	0.5
Value of Standard deviation for Enr(y) to be less than:	0.01
Expected average enrollment proportion:	0.2
Optimal Solution	$p_{ed} = p_{rd} = 0.3077$

Table 3: `linprog` optimization results

In this situation it appears to be in the school's best interest to make ED and RD acceptance rates equal to each other.

However, since r is proportional to $p_{ed} - p_{rd}$ (we can temporarily put 1 as our constant), we can create a new set of equations:

$$Mean(Enr(y)) = (p_{ed} - p_{rd}) \cdot p_{ed} + (1 - p_{ed} + p_{rd}) \cdot p_{rd} \cdot y_0$$
$$Var(Enr(y)) = (1 - p_{ed} + p_{rd})^2 \cdot {p_{rd}}^2 \cdot {\sigma_y}^2$$
$$Sd(Enr(y)) = (1 - p_{ed} + p_{rd}) \cdot p_{rd} \cdot \sigma_y$$

Optimizing this could provide different results.

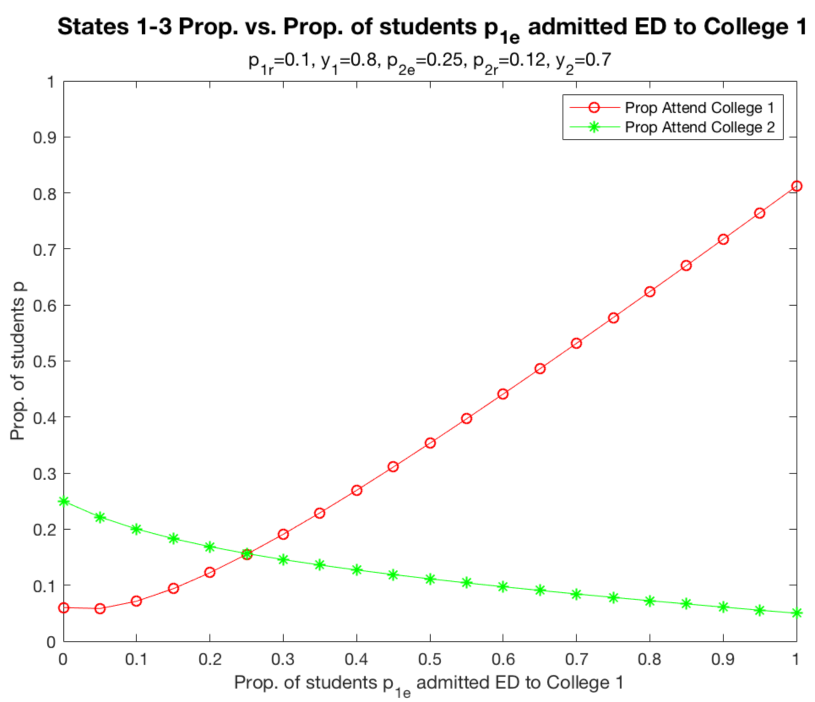
3.2 Two College Model

We can model this Two College Model with the following statistics for the admissions/yield rates:
$p_{1e} = 0.2,\ p_{1r} = 0.1,\ y_1 = 0.8,\ p_{2e} = 0.25,\ p_{2r} = 0.12,\ y_2 = 0.7$

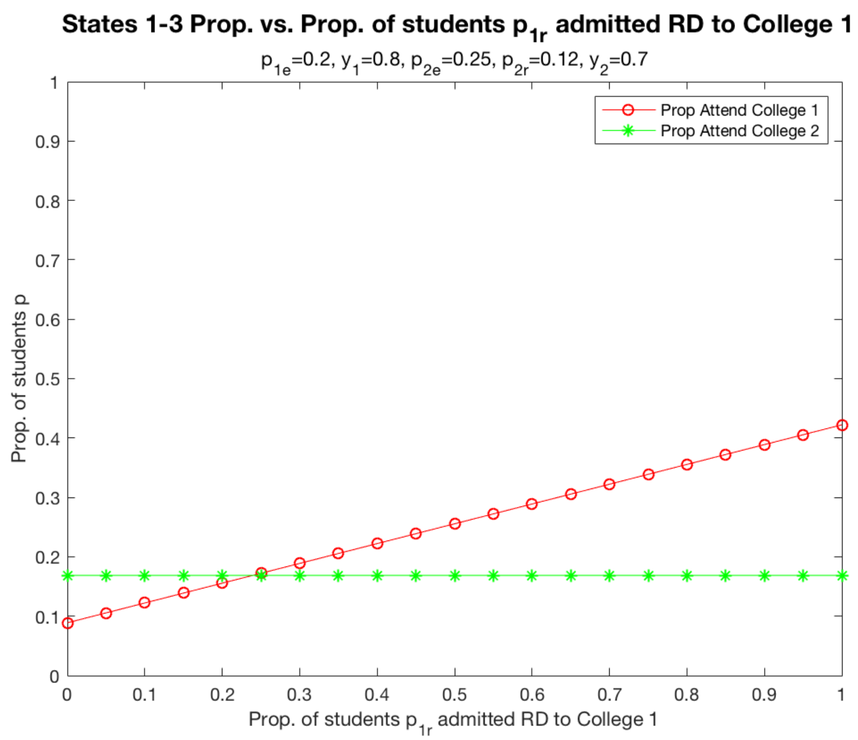
Proportion of original group attending via ED for College 1:	0.0889
Proportion of original group attending via ED for College 2:	0.1389
Proportion of original group attending via RD for College 1:	0.0333
Proportion of original group attending via RD for College 2:	0.0299
Proportion of original group attending neither college:	0.7090
Proportion of original group attending College 1	0.1222
Proportion of original group attending College 2	0.1688

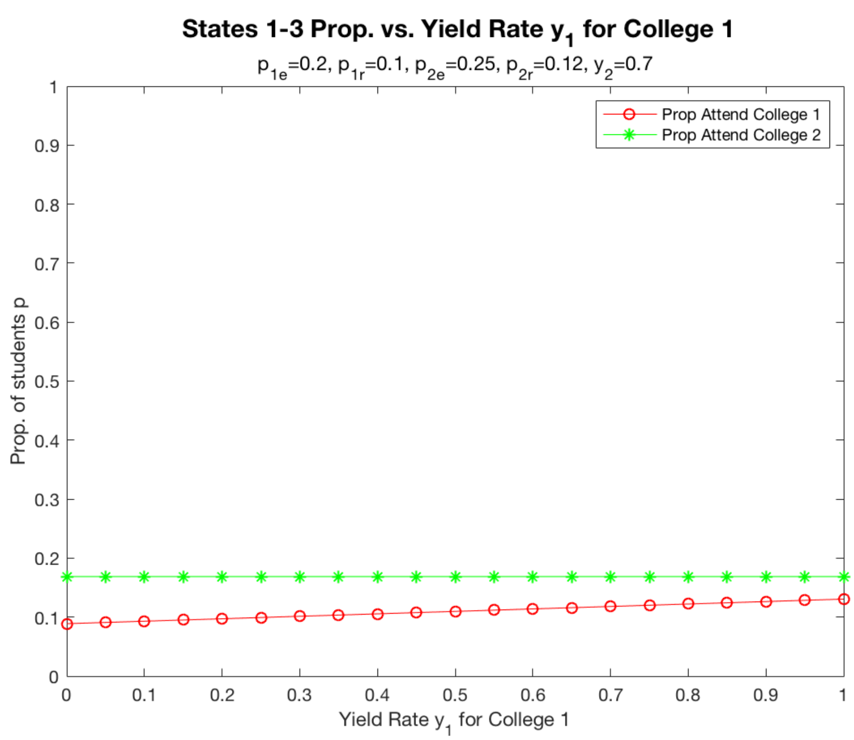
Table 4: Two College Model results

We can model the effect of $p_{1e},\ p_{1r},\ y_1$ on the final proportions by varying these values (while keeping the others constant) and graphing:



As the proportion of students p_{1e} admitted ED to College 1 increases, there is a higher proportion of people who will get in ED and attend College 1. Fewer people will get rejected ED from College 1, and fewer will turn to applying RD to College 2, so College 2's final attendance will become lower.

<table>
<tr>
<td>


</td>
<td>As the proportion of students p_{1r} admitted ED to College 1 increases so does the number of students attending College 1. The higher the admission rate, the more students attend (despite a constant yield rate). College 2's proportion of attendance is unchanged as students who apply RD to College 1 will never again apply to College 2 in the Markov Chain model.</td>
</tr>
<tr>
<td>


</td>
<td>As the yield rate y_1 at College 1 increases, more students attend College 1. The higher the RD yield rate, the more students attend (despite a constant RD admissions rate). College 2's proportion of attendance is unchanged as students who apply RD to College 1 will never again apply to College 2 in the Markov Chain model.</td>
</tr>
</table>

The institutions can refer to this model and adjust various admissions rates until an optimal balance of applicants attending is attained.

Section 4: Discussion of Results

As mentioned in Section 3, we overlooked a few exceptions in the Markov Chain system: defers, ED withdrawals, and rescindments—as examples. As such, our model with Cornell University's admissions were slightly off. The assumption of the proportion at which students apply either ED or RD was an educated guess; however, it does model the "hype" surrounding an ED program at a specific college.

Also, fortunately, our parameter adjustments' results have reflected real-life situations accurately. In the first model, increasing admissions rates increased enrollment at schools. An increase in ED admissions rates lead to fewer people in the original pool not attending the college; similarly, increasing RD admissions rates has an increase of RD attendance; however, the proportion of students not attending the college is relatively constant, likely due to the yield rate making a drastic change in that statistic. A higher yield rate pulled more RD students to attend, and as expected, left ED admissions alone.

The consideration of the standard deviation of the expected enrollment was based on the heavy assumption that the yield rate would follow a normal distribution—of course, risky, but also a necessary evil in order to create any result. Also, the `linprog` linear programming solver was used to only analyze a fixed rate of applicants applying to ED with rate r; the consideration with r being proportional to other factors was not considered with a limited MATLAB implementation with non-linear constraints.

With the two college system, we made a set of assumptions that all the students in the pool would apply to either college for ED, and switch to the other college for RD if rejected in ED by the first college. Also, we assumed students choosing to apply either school ED would be mostly out of how likely their success would be—if one college was more likely to accept students ED, then more students would apply there. That is a realistic explanation. Of course, the mathematical reasoning was a little uncomfortably arbitrary, though there is no easy way of quantifying this same kind of "hype" found in the first model.

Increasing ED success in one college would boost attendance at that college and take students from attending the other college; this matches with colleges competing over the same pool of qualified applicants and the ED program's attractiveness. Increasing RD success leaves no impact on the other college's attendance, because at that point no students will flow back to the college they were denied in the ED round. In real life, however, there might be a direct effect from the RD admissions, so a new consideration in the model would be needed. Yield, of course, has a very similar effect, as it did in the first Markov Chain system as well.

Section 5: Conclusion

The college admissions process is daunting to the students passing the gates; but it is often more stressful for the actual gatekeepers themselves—the admissions officers. In this project we have hoped to address certain ideals in the college admissions process—more specifically, the concerns regarding yield and Early Decision processes. We modeled the effects of different factors in the end results of college admissions and how colleges would potentially act upon these effects.

Hopefully in the future we can analyze the two college model more thoroughly. However, before analysing our two college model we must make some further assumptions:

1) The competitive pool (eg. a academically strong private school, magnet school, competitive public school) is small enough that colleges will not attempt to limit acceptances in favor of keeping admission rate down.
2) The two colleges wish to make the difference in proportion of students attending each college be as small as possible.

Therefore, for our second model, we want to minimize the difference between the student attending populations. We can keep this in mind.

By modeling college admissions with a Markov Chain, we oversimplify a complex social phenomenon; the purpose of this study is not to find a "hack" for either the student or the admissions officer, but rather to create generalized observations about how the system operates.

References

[1] Avery, Christopher, and Jonathan Levin. "Early Admissions at Selective Colleges." The American Economic Review 100, no. 5 (2010): 2125-156. http://www.jstor.org/stable/41038757.

[2] Fallows, James. "The Early-Decision Racket." *The Atlantic*. September 2001.
Accessed August 11, 2018.
https://www.theatlantic.com/magazine/archive/2001/09/the-early-decision-racket/302280/

[3] Schmidt, Samantha. "UC Irvine faces criticism after rescinding 499 admissions offers two months before start of school." *The Washington Post*. July 31, 2017.
Accessed August 11, 2018.
https://www.washingtonpost.com/news/morning-mix/wp/2017/07/31/u-c-irvine-faces-criticism-after-rescinding-499-admissions-offers-two-months-before-start-of-school/?utm_term=.b70203b056fb

[4] Krantz, Laura. "For small, private colleges, fewer students means more worries." *The Boston Globe*. March 31, 2018. Accessed August 11, 2018.
https://www.bostonglobe.com/metro/2018/03/31/for-small-private-colleges-fewer-students-means-more-worries/1jjd8ZFusBt3kGjHOcpIqM/story.html

[5] "Common Data Set 2017-2018" *Cornell University*.
Accessed August 11, 2018.
http://irp.dpb.cornell.edu/wp-content/uploads/2018/05/CDS_2017-2018-v5.pdf

[6] Chang, L. (2006), Applying data mining to predict college admissions yield: A case study. New Directions for Institutional Research, 2006: 53-68. doi:10.1002/ir.187

MATLAB Code

Attached in separate files.