Mathematical Modeling and Consulting

your figures & tables need to ~~be~~ blend in your text, e.g. don't just put them in one section ~~not just~~

See also WMA. Ch 5 & ch. 6

Sponsor

Blue Jays Unlimited

Final Report

Modeling and Simulating Fan Participation at Large Scale Sporting Events

Team Members

Ahmed Aly, Department of Biomedical Engineering
aaly2@jhu.edu

Steven Su, Department of Biomedical Engineering
ssu7@jhu.edu

Danni Tang, Department of Biomedical Engineering
danni@jhu.edu

Academic Mentor

Dr. N. H. Lee, Applied Mathematics and Statistics
nhlee@jhu.edu

Date: Last Compiled on December 8, 2012

This project is fictionally supported by Blue Jays Unlimited.

Abstract

Having a loud and supportive home crowd is the ultimate home team advantage for any sports team. This is especially true for collegiate sports. In this work we developed a simple stochastic model for modeling cheering in crowds at large sporting events. Our model accounts for differences in innate support level for a sports team between fans as well as the number and length of time surrounding fans have been cheering. It uses these factors to predict whether a single fan will start to cheer and then simulates over time how cheering in a crowd progresses. The results from the model are promising: cheering in a crowd can either increase greatly or increase slightly, depending on how the parameters of the model are varied.

Acknowledgments

We would like to thank Blue Jays Unlimited for fictionally sponsoring us. In addition, we appreciate the unwavering support and patient guidance our academic mentor, Dr. Nam H. Lee, has given us. We would also like to thank our classmates for their support.

Contents

List of Figures

List of Tables

Introduction

A loud and supportive home crowd is the ultimate home team advantage for any sports team. Past research has shown that the home crowd advantage can have a profound effect on deciding who wins a sporting event [1]. Across all major professional sporting leagues in the United States, the home team at any sporting event wins approximately 60% of the time, a statistically significantly higher winning percentage then the expected 50% [1]. The home crowd advantage is even more profound in college athletics. Studies have shown that in certain collegiate sports, the home team wins up to 66% of the games played [2].

In addition to helping the home team win, a loud and supportive home crowd can also improve the ambiance of a sporting event. Whether fans are chanting the school fight song, waving a rally towel, doing the wave, or clapping in general applause, the fans are not only showing support for the home team but also enhancing the general atmosphere of the sporting event at the same time. Note that all of these actions will hence be collectively termed "cheering." For this and previously stated reasons, it is evident that cheering is essential at collegiate sporting events. Cheering improves the collegiate sporting experience for both the athletic teams as well as the fans.

The Johns Hopkins University, located in Baltimore, MD, has a proud athletic tradition to go along with its long standing reputation for strong academics. The Johns Hopkins University Blue Jays have amassed 47 national championships and 187 conference titles [3]. The Blue Jays have excelled at many sports including Men's Lacrosse, Men's Swimming, Men's Football, and Men's Baseball. The Men's Lacrosse team has won 44 national championships, most recently in 2007, while the Men's Swimming team won 32 conference championships, including an astounding streak of 28 consecutive conference championships from 1971-1998 [3]. More recently, the Men's Football and Baseball teams were each conference champions for three consecutive years from 2009 through 2011 [3]. As with any collegiate athletic program, The Johns Hopkins University and its supporters are always looking for ways to improve its athletic teams' performances so that they may continue in their winning ways.

Background

Blue Jays Unlimited (BJU), established in 1995, is a volunteer group of alumni, friends and staff dedicated to supporting and promoting Johns Hopkins athletics [4]. BJU is the official booster club for Johns Hopkins athletics and has more than 3000 active members, who have raised more than $4 million in funds to improve the Johns Hopkins athletic experience for both student athletes and fans alike [4]. These funds provide money for capital projects as well as scholarship and operational endowments [4]. Past BJU projects have included renovations of the Newton H. White Athletic Center as well as recognition banners for championship teams [4].

BJU is present at nearly all major Johns Hopkins sporting events to encourage fans to support their Blue Jays in a vociferous and family-friendly manner to propel their Hopkins' teams to victory. It is their goal to provide Johns Hopkins' athletic teams with a spirited home crowd to cheer them on. As such, BJU is interested in maximizing the amount of fan participation in cheering at sporting events held on Homewood Field at the Homewood campus of The Johns Hopkins University. More specifically, they believe that they can increase fan participation in cheering events by strategically placing "cheer starters" in the home crowd to encourage other fans to participate in cheering. Cheer starters are student volunteers who lead and urge other fans around them to cheer.

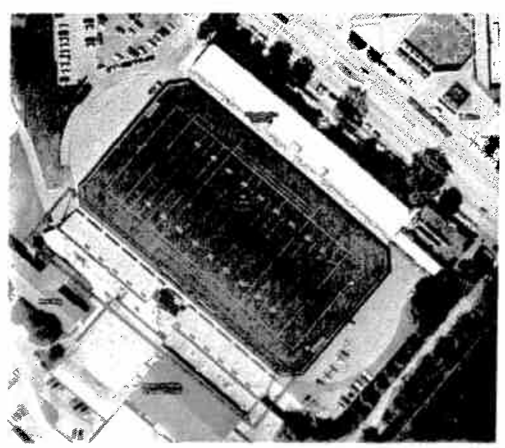


Figure 1: Homewood Field located on the Homewood campus of The Johns Hopkins University in Baltimore, MD. The bleachers in the lower left corner outlined in red are traditionally reserved for home team fans.

Consider the satellite image of Homewood Field in Figure 1, courtesy of Google Maps. Homewood Field's capacity is approximately 8500 spectators [5]. The long rectangular section of bleachers outlined in red in the lower left of portion of the image seats approximately 4000 fans and is traditionally reserved for Blue Jays' fans. For nearly all major Hopkins' sporting events, these home team bleachers are filled to capacity. As such, BJU is specifically interested in maximizing fan cheering in these home team bleachers.

Problem Statement and Objectives

Problem Statement

BJU wants to know if cheer starters can actually increase cheering and also wants a simple model of fan participation in cheering in the home team bleachers on Homewood Field.

~~Objectives~~ ok.

Our task is to provide BJU with a simple model of fan participation in cheering at Homewood Field in the home team bleachers as well as simulation results from the model which determine if their belief about cheer starters is accurate. If cheer starters are found to be effective, we will attempt to provide BJU with more details about the quantity ~~and location~~ at which cheer starters should be placed in order to maximize cheering, time permitting.

your work is about building a model for cheering. ok.

this is the final report?

Analysis

Because this model is a simulation, it is imperative to clearly define simplifications and assumptions we have made in the model.

Simplifications and Assumptions

- The willingness of a fan in a crowd to cheer depends on the number of people cheering around the given fan as well as how long the surrounding people have been cheering. The greater the number of people surrounding a fan who are cheering, and the longer the surrounding people have been cheering, the more likely that fan will start to cheer.
- The innate support level a fan has for the team also influences the willingness of that fan to start cheering.
- Once a fan starts to cheer, he/she continues cheering until the end of the simulation.
- The performance of the sports team does NOT influence cheering.

Computational Simulation

Using MATLAB, we start by generating an arbitrary sized n x m matrix, X, to represent a nm sized crowd. The number of rows, n, and columns, m, can be changed based on the user's liking. However, for our purposes we are choosing to make X a long rectangular matrix to imitate the physical dimensions of the home team bleachers on Homewood Field. The matrix X represents a crowd and each element in it corresponds to a specific fan in the crowd. For example X_{ij} represents fan ij. Each element in X is filled with the corresponding fan's innate support level for the team. Each fan's innate support level was generated by sampling from a normal random distribution with a mean of 10 and a standard deviation of 1 as shown in (1).

$$X_{ij} \sim Norm(10, 1), \ \forall \ i \in [1, n], \ j \in [1, m] \tag{1}$$

a bit disingenuous since no attempt to justify ~~numerically~~ is made... or ~~it's~~ it is a claim that you wish to establish in this paragraph

A normal distribution fits the distribution of innate support levels that a crowd of fans has very well. The majority of fans who attend sporting events are average or close to average fans; they will cheer if others around them cheer for a long enough time [6]. A smaller proportion are super fans who will cheer regardless of what anyone else around them is doing [6]. A smaller proportion of attendees may also be less supportive of the team; they require a greater amount of encouragement to start cheering than the average fan [6]. This is well represented in a normal distribution where most of the distribution is centered around the mean and an increasingly smaller proportion of the distribution is found moving away from the mean in both directions [6]. In this model the mean was set to

10 and the standard deviation was set to 1 to ensure that it was highly unlikely that a fan was assigned a negative innate support level. (i.e. it is very unlikely to sample a number from the normal distribution that is more than 10 standard deviations from the mean.) This makes the math much easier when considering the dependence on the surrounding people, as it will be shown later.

some time means transpose

Next, we set an initial threshold, T_{init}. The initial threshold is used to determine which of the fans in the matrix are *initially* cheering. Another n x m matrix, X', was then created. Each element in X' is initially assigned a value of 1, if the corresponding element in X (i.e. the element in X with the same row and column indices) had a value which was greater than or equal to the initial threshold, otherwise the element is given a value of 0. See (2). X' is the matrix used to keep track of who is cheering. A value of 1 in X' means that given fan is cheering, while a value of 0 means that given fan is not cheering.

period

$$X'_{ij} = 1 \; if \; X_{ij} \geq T_{init}, \; X'_{ij} = 0 \; if \; X_{ij} < T_{init} \qquad (2)$$

Any fan with an innate support level greater than 1 standard deviation above the mean is far above the average fan and can be considered a super fan who would be initially cheering. As such, T_{init} was appropriately set to equal 11, which is 1 standard deviation above the mean of the normal distribution used to generate the innate support levels.

As previously stated, whether a fan starts to cheer depends on how many directly adjacent fans around the given fan are cheering, as well as the length of time those fans have been cheering. S is a n x m matrix which stores how many people surrounding a given fan are cheering at a given time. Each element in S has corresponding elements with the same row and column indices in both X and X', all of which store different model values for the same individual fan. We define a round, r, to be the passing of an arbitrary time interval (approximately 3-5 seconds, in this case). We then create a fourth n x m matrix, Y, and compute the individual elements of Y according to (3).

$$Y_{ij} = X_{ij}(S_{ij}) + r, \; \forall \; i \in [1, n], \; j \in [1, m] \qquad (3)$$

Again, each element in Y represents an individual fan and has corresponding elements in X, X', and S, with the same row and column indices. We then compare each of the elements in Y to an absolute threshold, $T_{absolute}$. If an individual's score in Y is greater than or equal to the absolute threshold, the individual will begin to cheer, and we set the corresponding element in X' equal to 1. This is shown in (4). If the individual's score in Y is less than the absolute threshold, we do nothing to the corresponding element in X' (i.e. it remains 0).

$$X'_{ij} = 1 \; if \; Y_{ij} \geq T_{absolute} \qquad (4)$$

We repeat this process for each round (i.e. compute Y, update X' based on the new Y, and then update S for the next round based on the new X') until we reach the desired number of rounds, R. During each round, we take snapshots of the crowds by storing X' for that round. By doing so, we can compare X' between the rounds and also compute the percentage of fans who are cheering in each round. It is important to note, by computing Y as shown in (3), the likelihood of a fan starting to cheer increases with the number of surrounding fans who are cheering (i.e. as S_{ij} increases) and also increases with the length of time those fans have been cheering (i.e. as r increases).

To test the effect of the number of cheer starters, we randomly place CS number of cheer starters in the crowd. To implement the CS cheer starters into the model, when we initialize X' we randomly choose CS distinct row and column indices in X' and assign those elements to equal 1. By doing this, we insure that the CS cheer starters are initially cheering and remain cheering for the entire simulation. The crowd simulation was then run and the final percentage of cheering fans after 10 rounds was computed as previously described. We repeated this 1000 times for the given CS value and for each of the 1000 trials we recorded the final percentage of cheering fans after the 10 rounds. We then averaged these percentages over all 1000 trials to come up with the average final percentage of cheering fans for that given CS value. We repeated this Monte Carlo procedure for all CS values in the range of $1 \leq CS \leq 50$. The average final percentage of cheering fans for each CS value was then compared to the average final percentage of cheering fans when there were zero cheer starters, using a simple t-test.

Final Parameter Values

Due to computational limits, we could not simulate a crowd of 4000 fans. Instead we downsized our crowd size to 2000 fans and choose to simulate a crowd of 20 by 100 fans. The matrix produced by simulating 20 by 100 fans is long and rectangular and thus similar in shape to the home team bleachers at Homewood Field. We decided to do 10 rounds per simulation.

For the threshold values, we decided to set our initial threshold at 11 and the absolute threshold at 46. We chose these threshold values after extensive testing where we tried different combinations of initial threshold and absolute threshold values. We found that the values we selected were appropriate. When the initial threshold was set below 11, we ~~thought~~ that there were too many fans that were initially cheering. For example as seen in Figure 1, when $T_{init} = 10$ and $T_{absolute} = 46$, approximately 50% of the crowd is initially cheering. When the initial threshold was greater than 11, to few fans were initially cheering and there was no increase in cheering. This can be seen in Table 1; when $T_{init} = 12$ and $T_{absolute} = 46$, only 2.5% of the crowd is initially cheering and there is no increase in cheering. When the absolute threshold was lower than 46, we felt that too high of a percentage of fans ended up cheering to be realistic. For example in Table 1 when $T_{init} = 11$ and $T_{absolute} = 40$, 92.70% of the crowd ends up cheering. When ~~the absolute~~ threshold was higher than 46, not enough fans ended up cheering to be realistic. In fact if the absolute threshold was set high enough, the simulation tended to stall, meaning that there was little to no increase in cheering after a certain round, and we felt this was also unrealistic. This stalling can be ~~seen in Table 1 when~~ $\cancel{T_{init} = 11}$ ~~and~~ $\cancel{T_{absolute}} = 65$. When $T_{init} = \cancel{11}$ ~~and~~ $\cancel{T_{absolute}} = 46$, initially 15.5% of the crowd is cheering and at the end of the simulation 63.8% of the crowd is cheering. We felt these values to be appropriate.

The final parameter values we used are as follows:

- Rows, $n = 20$
- Columns, $m = 100$

- Initial Threshold, $T_{init} = 11$
- Absolute Threshold, $T_{absolute} = 46$
- Rounds, $R = 10$
- Number of Cheer Starters, CS (Experimental Variable)

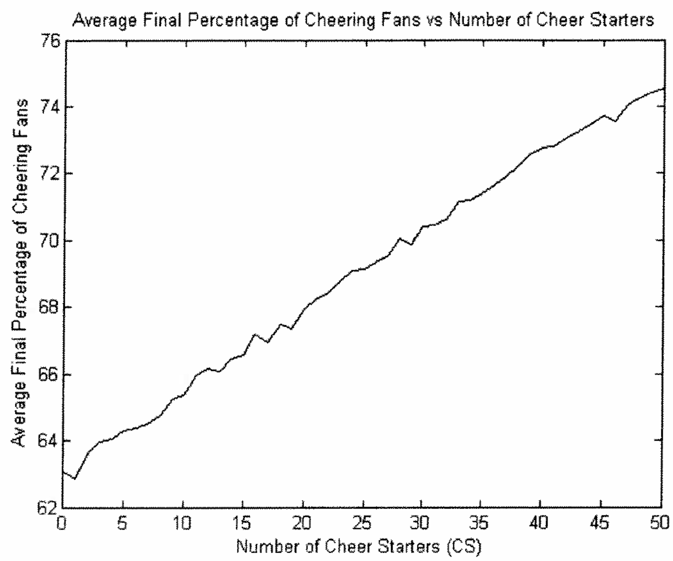
~~Abd~~ why all alone?

Round	1	2	3	4	5	6	7	8	9	10
$T_{init} = 10,\ T_{abs} = 46$	49.7	74.2	87.15	93.8	97.4	98.85	99.95	99.95	99.95	99.95
$T_{init} = 12,\ T_{abs} = 46$	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
$T_{init} = 11,\ T_{abs} = 40$	14.85	19.85	26.2	35.00	46.00	57.70	69.25	79.55	87.45	92.70
$T_{init} = 11,\ T_{abs} = 65$	15.15	15.25	15.40	15.40	15.45	15.5	15.55	15.55	15.55	15.55
$T_{init} = 11,\ T_{abs} = 46$	15.5	16.65	18.05	20	23.15	27.8	34	42.2	52.2	63.8

Table 1: Percent of cheering crowd over rounds for some combinations of T_{init} and $T_{absolute}$ values.

why whitespace?

Results



??

Figure 2: Average final percentage of cheering fans over 1000 trials for various CS values.

Number of Cheer Starters	0	10	20	30	39	50
Average Final Percentage of Cheering Fans	63.069	65.346	67.919	70.409	72.533	74.544
P-value	0.5	0.34413	0.19637	0.097951	0.047687	0.021585

Table 2: Average final percentage of cheering fans for various CS values.

Not explained in text?

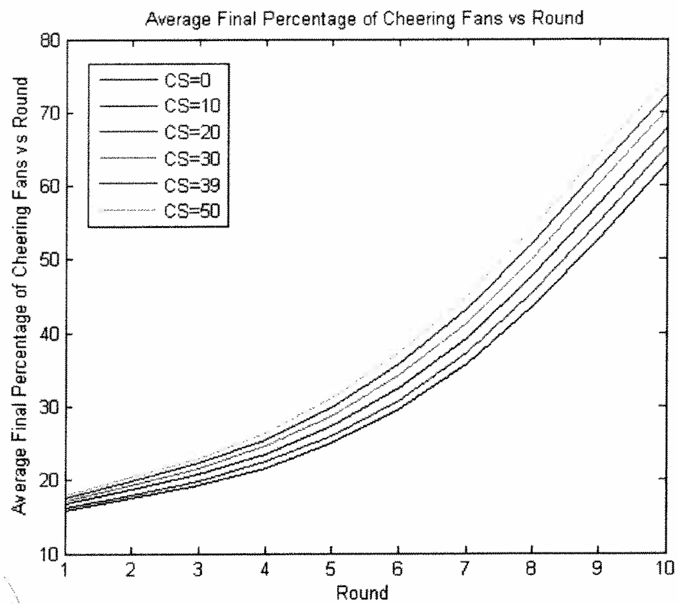


Figure 3: The percentage of cheering fans over time for various CS values.

not used in text?

Round	1	3	7	10
CS=0	15.875	19.285	35.67	63.069
CS=10	16.266	19.936	37.307	65.346
CS=20	16.686	20.717	39.246	67.919
CS=30	17.13	21.53	41.284	70.409
CS=39	17.519	22.261	43.088	72.533
CS=50	17.908	23.002	44.936	74.544

Table 3: The percentage of cheering fans over time for various CS values.

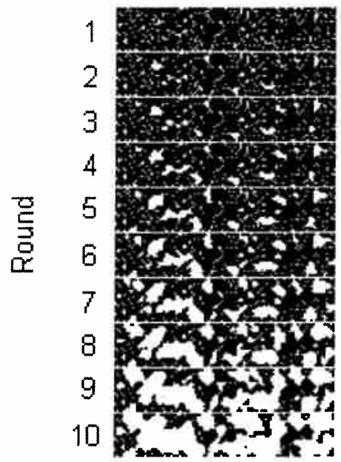


Figure 4: Cheering over time when $CS = 0$. White indicates cheering.

How should one interpret this?

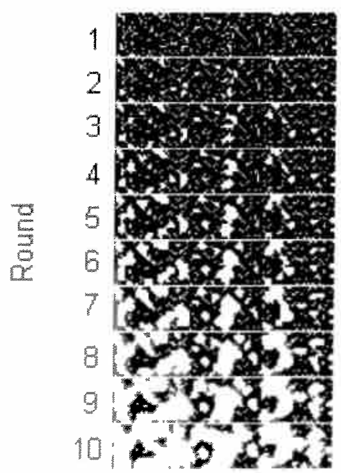


Figure 5: Cheering over time when $CS = 39$. White indicates cheering.

See Appendix A of WMA (Chicago) pg 407 – 416.

Discussion

By testing CS values between 1 and 50, we found that when $CS \geq 39$, there is a statistically significant increase in cheering after 10 rounds have passed ($p = 0.047$). Another observation we made was that if the number of cheer starters is increased further, the average final percentage of cheering fans also increases and the p-value of the t-test decreased. This was an expected result, since if there are more cheer starters in the crowd, more people should end up cheering. Thus this is consistent with our model.

o.k.

Figure 2 shows a plot of the average final percentage of cheering fans versus the number of cheer starters. The data on this plot is averaged over the 1000 trials for each CS value. We observe a positive linear relationship between the average final percentage of cheering fans and the number of cheer starters. As the number of cheer starters increases, so does the average final percentage of cheering fans. Some actual values from the plot in Figure 2 are listed in Table 2. When there are 0 cheer starters, approximately 63% of the crowd ends up cheering. When there are 39 cheer starters, around 73% of the crowd ends up cheering. As mentioned before, 39 is the first CS that produces a statistically significant increase in cheering where the p-value is less than 5%.

Figure 3 shows a plot of the percent of the crowd cheering in each round for several CS values. Note that the data in this plot is also averaged across 1000 trials for each CS value. Based on this plot we can see that all the trajectories of the plots are similar in shape, regardless of the CS value, however trajectories with greater CS values have a steeper slope. This indicates that the more cheer starters that are in the crowd, the greater the rate of increase of cheering.

Table 3 contains actual values from the plot in Figure 3. At round 1, when $CS = 0$, approximately 16% of the crowd is initially cheering. After 10 rounds this percentage grows to around 63% as mentioned before. When $CS = 39$, approximately 17.5% of the crowd is initially cheering and that this percentage grows to around 72% after 10 rounds. Thus, after 10 rounds there is approximately a 10% difference in the final percentage of cheering fans between when there are 0 cheer starters and when there are 39 cheer starters. This difference can be seen in Figure 4 and Figure 5 which show an approximate 10% difference at round 10.

this is a numerical summary of your experimental result. Explain (and Discuss) what ~~do~~ it mean ~~bad~~ ~~good~~ for general audience.

Conclusion

The goals of this project have been accomplished and the deliverables have produced in a timely fashion. The following list summarizes our goals and objectives in a clear and concise manner:

- Provide BJU with a simple model of fan participation in cheering at Homewood Field,
- Provide simulation results from the model which determine if cheer starters are effective in increasing cheering,
- If cheer starters are effective, and time permitting, we will provide BJU with details about the quantity and locations at which cheer starters should be placed in order to maximize cheering.

Deliverables

- MATLAB R2009b and R combination package with test scripts that can be used to reproduce our numerical and simulation test results
- Technical report and presentations summarizing the work
- Time permitting, a list of cheer starter setups which maximize cheering

All deliverables were completed, except for the last one due to time limitations.

After testing different parameters, we can safely conclude that this model is reasonably qualitalively accurate; it follows the general assumptions and observations. We believe that this model will be viable for the future, and will serve as a framework for future research into sociodynamic studies. Ideally, this model can be applied to almost all settings with audiences or crowds, and can be used influence the audience to actively acquiesce to an opinion or action.

Glossary

Cheering. Any action performed by a fan which shows support for a sports team. Examples include chanting the school fight song, waving a rally towel, doing the wave, or clapping in general applause.

Cheer Starter. A student volunteer who leads and urges other fans around him/her to cheer.

Abbreviations

BJU. Blue Jays Unlimited

Selected Bibliography Including Cited Works

[1] Jeremy P. Jamieson. The Home Field Advantage in Athletics: A Meta-Analysis. *Journal of Applied Social Psychology*, 40(7):1819–1848, 2010.

[2] E.E. Snyder and D.A. Purdy. The Home Advantage in Collegiate Basketball. *Sociology of Sport Journal*, 2(4):352–356, 1985.

[3] Johns Hopkins Tradition. `http://www.hopkinssports.com/trads/conference-champs.html`. Accessed: 11/02/2012.

[4] Blue Jays Unlimited - Johns Hopkins Official Athletics Site. `http://www.hopkinssports.com/bluejays-unlimited/`. Accessed: 10/12/2012.

[5] Homewood Field. `http://en.wikipedia.org/wiki/Homewood\_Field`. Accessed: 10/12/2012.

[6] Roger Evertt. *Diffusion of Innovations*. Free Press, A Division of Simon and Schuster Inc., 5th edition, 2003.