



## NCAA Basketball Tournament Analysis for High School Mathematics

### **Dr. Adrian J Lee, Central Illinois Technology and Education Research Institute**

Dr. Adrian Lee received his Ph.D. in mechanical engineering from the University of Illinois at Urbana-Champaign in 2009, specializing in probability and risk analysis of aviation security systems. Dr. Lee served as a post-doctoral research engineer at Vishwamitra Research Institute, Center for Uncertain Systems: Tools for Optimization and Management, and is currently President of Central Illinois Technology and Education Research Institute. Dr. Lee also holds an adjunct academic position in the Department of Bioengineering at the University of Illinois at Chicago, and is a member of the Institute for Operations Research and Management Science (INFORMS), the Institute of Electrical and Electronics Engineers (IEEE), the Society for Industrial and Applied Mathematics (SIAM), and the American Society for Engineering Education (ASEE). His research interests include STEM education, probability and statistics, and stochastic optimization.

### **Prof. Sheldon H. Jacobson, University of Illinois, Urbana-Champaign**

Sheldon H. Jacobson is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. He has a B.Sc. and M.Sc. in Mathematics from McGill University and a Ph.D. in Operations Research from Cornell University. His research interests span theory and practice, covering decision-making under uncertainty and discrete optimization modeling and analysis, with applications in aviation security, health care, and sports. From 2012-2014, he was on leave from the University of Illinois, serving as the Program Director for Operations Research in the Division of Civil, Mechanical and Manufacturing Innovation at the National Science Foundation.

### **Mr. William A Cragoe, Sacred Heart-Griffin High School**

I am currently (as of 3/7/14) in my 8th year as a high school mathematics teacher. I recently started teaching a Statistics course and this will be the second year of using "Bracketodds" as a class project.

# NCAA Basketball Tournament Analysis for High School Mathematics

## Abstract

As the winter season steadily makes way for spring, basketball fever heats up when the sports media begin to headline various qualification scenarios for the annual NCAA men's basketball tournament. College basketball experts and sports analysts provide wisdom into how the tournament field might be seeded and which teams are anticipated to reach the coveted Final Four. The media hype preceding the tournament generates excitement and competitiveness amongst sports fans nationwide as each individual strives to predict the elusive perfect bracket. The popularity of this competition coupled with the uncertainty of buzzer-beating upsets provide a unique and interesting opportunity to learn how probability methods can be used to model and predict real life events. This paper outlines a week long instructional curriculum for high school math and engineering classes based on prior published academic research on a theoretical predictive model. The underlying concept is based on a sequence of Bernoulli trials, where a mathematical model captures the probability of a particular seeded team advancing in each round according to a geometric distribution. These basic concepts easily fit within the scope of high school probability and statistics, and when delivered in the days prior to tournament tip-off, the curriculum provides an excellent opportunity to inspire students into addressing real world problems through mathematical analysis.

## I. Introduction

Commonly referred to as “March Madness”, the NCAA men's basketball tournament fuels three weeks of excitement (and anguish) nationwide as fans root for their favorite collegiate teams to advance through each stage of the competition. Following a committee selection process and set of four initial play-in games, sixty four teams – seeded 1 through 16 in four separate regions – participate in a single elimination tournament format to determine who will be crowned national champion. The structure of such a competition, coupled with the immense national interest, makes it an ideal event for the creation of so-called “office pools”, where the general population attempts to predict which teams will advance by filling in brackets prior to the start of the tournament. Rather than basing these decisions on favorite teams, uniforms, or mascots, one can gain a better understanding of the likelihood of certain seeded teams advancing in each round based on the statistics associated with the tournament's prior historical results.

This work applies introductory level probability methods towards the analysis of the NCAA men's basketball tournament in an exciting week long instructional session for high school math and engineering classes. The application of this work is currently in its second year of instruction, where the theoretical material is derived from prior academic research published by Jacobson et al.<sup>1</sup>. During the week prior to Selection Sunday – the day teams are selected and seeded for the tournament by a committee of experts in the collegiate basketball community – students learn how the truncated geometric distribution can be used to model the likelihood of seeds advancing in each round. The results from the past 29 tournaments are used to validate the model based on a chi-squared goodness-of-fit test. Students learn how mathematics can be used to model uncertainty and gain a better understanding of the outcome of random events through a real world scenario. A combination of lecture slides and computational analysis using Microsoft

Excel allow the students to learn about the underlying probability concepts, and then apply them through computer simulation exercises. In-class and homework assignments provide indications of how well the students understand the underlying concepts.

Presently, a few courses teaching the concepts of “Bracketology” are offered at the university level. The University of Cincinnati’s College of Business offers a course teaching a statistical approach to predicting the tournament winners<sup>2</sup>. This approach is based on specific regular season statistics associated with each team participating in the tournament. Then, the most likely bracket outcome is obtained by simulating over several thousand combinations of potential matchups. The College of Professional and Liberal Studies at St. Joseph University also offers a “Fundamentals of Bracketology” course online, which focuses more towards the structure and history of the NCAA tournament<sup>3</sup>. Using an entirely different approach, this paper outlines the deployment of a shorter, week long instructional curriculum designed for a high school audience. Its predictive model is based on the seeding format for the tournament, where individual team performance is encapsulated through the NCAA selection committee’s seeding assignments.

The paper is organized as follows: Section II discusses the classroom instructional material presented during the first three days of the five-day curriculum, including the tournament history, the theoretical model, and an exercise in past tournament analysis. Section III presents an opportunity to use this effort to promote future enrollment in math and engineering by offering a school wide tournament challenge. The data gathered from the brackets submitted for this challenge are then analyzed by the class during the last two days of instruction. Section IV discusses potential engineering applications which may benefit from the use of probability-based models, along with proposing future directions in expanding the instructional material. Section V ends with concluding remarks.

## **II. Classroom Instructional Material**

The instructional curriculum covers a five day period using a combination of lectures and in-class activities, held (ideally) Monday through Thursday following Selection Sunday, with the fifth day occurring upon the completion of the tournament. Day 1 begins the school week with a brief historical background of the NCAA basketball tournament and its competition format. Day 2 presents the mathematical theory used to derive the probability-based model. Day 3 allows the students to use the model to investigate the likelihood of past tournament results. Then on Day 4, the commencement of games in the round of 64 teams, the students conduct a pre-tournament analysis of the brackets submitted in a school wide tournament challenge. Following the completion of the NCAA tournament, the curriculum concludes on Day 5 with a post-tournament analysis regarding the likelihood of the eventual champion. The following describes the contents of each day in further detail.

### Day 1: Tournament Background

A quick poll of the 25 junior and senior students (14 female and 11 male) enduring the pilot study of this curriculum indicated that over one-half had little familiarity with the NCAA basketball tournament, while a few eagerly offered their round by round predictions. Consequently, the entire first day of instruction is dedicated towards covering a brief historical

background of the tournament, the NCAA committee selection and seeding process, and the tournament's current format of six single elimination rounds with four play-in games. This background creates a foundation for building the mathematical model, and to help the students realize how a model can be constructed using underlying characteristics of a physical event.

A second important element of this initial day is to reference and draw comparisons to existing predictive models and ranking systems, such as Rating Percentage Index (RPI), Sagarin, Massey, and Pomeroy, for example. These rating systems are based on factors including the outcomes of regular season games, score differential, and strength of schedule. True, the use of these systems can help influence one's decision making process. However, a more simplistic (in terms of implementation), numerical, and probabilistic method can look rather attractive in comparison. The purpose of the curriculum taught in this week long course is to demonstrate that a mathematical approach based on the tournament *seeds* rather than individual team performance can yield a viable solution to understanding the likelihood of unexpected results.

### Day 2: The Math behind the Numbers

The purpose of this paper is to focus on classroom implementation rather than the mathematical theory. Therefore, the reader is referred to the work by Jacobson et al.<sup>1</sup> for an in-depth explanation of the underlying theory. However, to not disappoint those in search of a symbolic expression, the following theorem is key to modeling the advancement of seeds in the tournament by means of a necessary and sufficient condition for a geometric random variable<sup>4</sup>:

Let  $X_1, X_2, \dots$  be an arbitrary sequence of Bernoulli trials. Let  $Z$  be the number of these Bernoulli trials until the first success. Then  $Z$  is a geometric random variable with probability  $p$  if and only if

$$P\{X_i=1 \mid \sum_{h=1,2,\dots,i-1} X_h = 0\} = p \text{ for all } i = 1,2,\dots$$

This theorem states that if the probability of the  $i^{\text{th}}$  trial being a success, given that the previous  $i-1$  trials failed is equal to  $p$ , and that value is constant regardless of the number of trials conducted, then the number of trials until that first success is a geometric random variable. This statement can be applied to any particular round of the tournament by analyzing the probability that the  $i^{\text{th}}$  remaining seed will advance to the next round, given that the  $i-1$  higher ranking seeds did not. Statistical analysis based on the method of moments using data from the past 29 national tournaments (i.e., those with 64 team fields) indicates that this geometric model does indeed fit well, especially for the latter rounds of the tournament.

For the level of high school probability and statistics classes, teachable concepts including discrete random variables, Bernoulli trials, binomial and geometric distributions, expected value, and the method of moments are easily within reason. One aspect of the analysis that expands upon these tools involves creating a truncated distribution. Considering there are a finite number of teams competing in each round, the geometric distribution must be truncated to meet the condition that all probabilities sum to one. For example, for the set of seeds  $\{1,8,9,16\}$  potentially advancing to the third round (a.k.a. Sweet Sixteen), the probability of Seed 1 advancing, using the standard geometric distribution, is  $p$ , Seed 8 is  $p(1-p)$ , Seed 9 is  $p(1-p)^2$ , and Seed 16 is  $p(1-p)^3$ . However, for these four probability values to sum to one, each must first be multiplied by the term  $1/(1-(1-p)^4)$ , thereby truncating the distribution over the four outcomes.

### Day 3: Prior Tournament Analysis

The truncated geometric distribution was shown in Day 2 to fit the historical data well, statistically, using the method of moments. The continuation into Day 3 then focuses on using this validated model to analyze the likelihood of past tournament outcomes. Questions can be posed such as:

- How often has each seed advanced to each round, compared to what are the expected number of times that seed should have appeared in that round?
- How often should we expect to see a certain combination of seeds (e.g., all No. 1 seeds) in the Final Four?
- What might the probability be that one or more lower seeded teams (11 through 16) make the Final Four?
- What is the most likely seed combination to occur in the Elite Eight, Final Four, the National Finalists, and as National Champion?

Moreover, one can use this model to draw comparisons between the likelihood of two seed combinations occurring, thereby answering the number one question: *Is it truly the best strategy to pick the better seeds?* This component of the curriculum allows the students to understand how a validated model can be used to provide answers to what might be critical decisions in a real world scenario. As opposed to merely guessing or following a biased selection, the mathematical model is shown to provide a quantifiable measure for which to base these decisions, reinforce or contradict intuition, and improve risk management.

A useful application of the truncated geometric distribution model is that it also serves as a good indication of which tournament seeds the public, as a whole, tend to favor in their bracket selections. Popular tournament challenges through ESPN and CBS Sports garner several million bracket predictions from hopeful fans across the nation (and abroad, but alas, they are not eligible to win!). True, many of these are based on favorite collegiate teams winning it all, while a few brave souls decide this is the year all No. 16 seeds make the Final Four. These challenges also allow people to submit several brackets per user account, thereby generating non-independent variations of potential outcomes. For a large, diverse sample size, however, the computed probability of a seed reaching the Final Four can be compared against the percentage of brackets submitted with that seed chosen in that round. Likewise, the probability of a seed advancing to a particular round can be compared numerically to the probability that someone from a large population will also decide to select that seed. These types of comparisons help indicate how close the selection strategies based on human intuition compare to past tournament results. Note that this observation does not necessarily hold when considering the probability of selecting a team assigned that seed, as historically successful teams tend to receive larger public support, as shown by the following illustration.

Take for example No. 1 Kentucky versus No. 2 Kansas in the 2012 championship match. Of the 6.45 million ESPN brackets submitted, 2,263,950 chose eventual champion Kentucky to win (in comparison to 820,762 expected brackets based on the geometric model), while 403,125 chose Kansas to win (399,900 expected). Of the two other Final Four contenders, 309,600 chose No. 2 Ohio State (403,125 expected) and 51,600 chose No. 4 Louisville (97,072 expected). Apart from the public bias towards Kentucky as the overwhelming tournament favorite, the scaling of public support matches well with the likelihood of the remaining seeded teams winning the

competition (note: ESPN publically provides only limited amounts of data pertaining to the how many brackets chose a particular team, rather than all teams seeded No. 1, 2, etc.). Clearly, this type of analysis does not address the size of a team's fan base, nor does it distinguish among the four No. 1 seeds across the four regions (or No. 2, 3, ... seeds for that matter) – case in point where nearly 30% more brackets favored Kansas in comparison to the identically seeded Ohio State. Despite these shortcomings, it is important to stress during this Day 3 lecture that the basic model serves as a foundation for constructing more complex algorithms which account for additional factors. In the next section, the class is given the opportunity to further investigate this behavior by analyzing data obtained from a school wide tournament challenge.

### III. School Wide Implementation

In addition to directly involving the students being taught the material in this study, students and faculty at the high school who are not directly connected with the class are invited to participate in a school wide tournament challenge, where each person can submit a bracket at no cost in the hopes of winning token prizes, including school apparel, accessories, and even the school's beloved cafeteria cookies. This event helps promote the classroom activities in the hopes of motivating students to enroll in the elective class in the future. The number of submitted brackets along with the year-to-year class enrollment will, over time, indicate the effectiveness of promoting the class through this school wide tournament challenge.

The last two days of the instructional curriculum consist of a pre-tournament analysis of the school brackets, followed three weeks later by a post-tournament analysis comparing the actual tournament outcomes to their predicted likelihood of occurrence.

#### Day 4: Pre-tournament Bracket Analysis

In the initial year of establishing the school wide tournament challenge, each student or faculty member was permitted to submit a single completed bracket by school dismissal on Day 3 (i.e., the day before the first round of tournament games commence). This collection of brackets serves as a source of data for the students to tally the number of people selecting each seed to reach rounds of the Elite Eight, Final Four, finalists, and eventual champion. From the school's population of roughly 900 students and faculty, 143 participated in the school wide tournament challenge (~15% participation rate) in 2013. Participation by grade level included 30 seniors, 23 juniors, 30 sophomores, 31 freshmen, and 9 faculty members. Twenty "zero-hour" eighth grade students participated, who attend early morning math classes at the high school from several surrounding area junior high schools. Although this number is a far cry from the several millions of brackets submitted to ESPN and CBS Sports, it is a more manageable quantity to divide amongst the classroom students to quickly count the frequency of seeds appearing in each of the latter rounds.

Table 1 lists the number of times each seed was chosen in the set of school brackets as national champion. The students are then asked to take the probability of each seed becoming champion and compute the expected number of the 143 school brackets to have selected that seed (i.e.,  $E[\text{no. brackets}] = 143 * p_i$ ). A normalized prediction error between these two values,  $a$  and  $b$ , is computed using  $\delta_i = (a-b)^2/b$ , for each  $i = 1, 2, \dots, 16$  seed.

Table 1: Student Body Predictions

National Champion	Number of Student Selections, $a$	Expected Number of Brackets, $b$	Prediction Error, $\delta_i$
1 seed	78	72.8	0.37
2 seed	51	35.8	6.45
3 seed	5	17.6	9.02
4 seed	2	8.6	5.07
5 seed	3	4.2	0.34
6 seed	1	2.1	0.58
7 seed	1	1.0	0
8 seed	0	0.50	0.50
9 seed	1	0.25	2.25
10 seed	0	0.12	0.12
11 seed	1	0.06	14.73
12 seed	0	0.03	0.03
13 seed	0	0.01	0.01
14 seed	0	0.01	0.01
15 seed	0	<0.01	0
16 seed	0	<0.01	0
			$\chi^2 = \sum_i \delta_i = 39.48$

Using the results in Table 1, a chi-squared test with  $\chi^2 = 39.48$  and  $n = 15$  degrees of freedom (Excel CHIDIST( $\chi^2, n$ ) function) results in a statistical p-value of 0.00054. This low p-value indicates that the school population's intuition regarding their strategy of selecting the champion does not closely follow the historical likelihood of an eventual champion's seed. The lack of expected No.3 and No.4 seed choices, along with more than expected No.2 seed choices demonstrates that student population's intuition instead appears to follow the strategy of selecting the best seed to win.

In addition to this exercise, students are asked to complete a bracket themselves – as homework, choosing their favorite teams – and compute in class the odds against their seed combinations reaching the Elite Eight, Final Four, and championship rounds. They then pair up to compute the relative likelihood between their seed combinations to see who has the matchup most likely to occur. The concept of conditional probability is addressed by asking what the probability is of their championship seed to win, given their choice of Final Four teams.

#### Day 5: Post-tournament Bracket Analysis

The fifth and final day of instruction revisits the material by conducting a post-tournament analysis of the school's success as a whole in correctly picking the realized seed combinations in each round, as well as discussing publically disseminated information regarding the large-scale ESPN and CBS Sports tournament challenges.

Using a point scoring system equivalent to that used by ESPN, a maximum of 1920 possible points (320 pts per round) is attainable if each and every winning team is selected in each round (i.e., the perfect bracket)<sup>5</sup>. Figure 1 shows a histogram of the final points achieved by those

participating in the 2013 school wide challenge. Interestingly, the results do not clearly follow a single normal distribution. However, this single data set containing a limited number of brackets is not yet sufficient to draw any conclusions about what point total one can expect to achieve by predicting outcomes in this type of tournament format. A larger data set, such as those obtained by ESPN or CBS Sports, analyzed over several years might help indicate this characteristic.

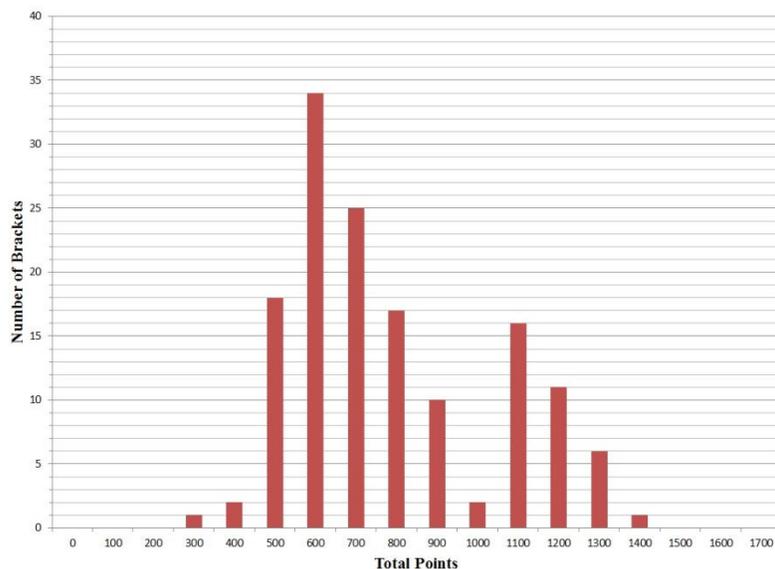


Figure 1: Distribution of School Bracket Point Totals

What Figure 1 is useful to illustrate, however, is the fact that the vast majority of hopeful participants clearly end up nowhere near topping the list of point earners. Moreover, the final point totals are not necessarily consistent with the individual's knowledge of the teams competing in the tournament. Consequently, students learn that while one bracket inevitably scores the highest, it is highly unlikely that their own bracket will achieve that total. Instead, the lesson being taught is that the use of mathematics can help improve one's chances of making correct decisions, but still cannot predict outcomes of a random event with 100% certainty. How did those following the path of the model's highest likelihood of final four seed combination  $\{1,1,2,3\}$  fair? After the 4<sup>th</sup> round, two of the sixteen students who chose  $\{1,1,2,3\}$  final four seed combinations were leading the eventual school champion ( $\{1,2,2,4\}$  seed choices) with scores of 800 and 770 versus 720 points. The pivotal moment came when the winning student correctly selected No. 1 Louisville vs No. 4 Michigan to appear in the championship match.

In addition to the final point totals, the number of school brackets containing the correct seed combinations in later rounds can also be counted. Table 2 lists the likelihood of the actual seed combination to appear in the Elite Eight, Final Four, finalists, and championship rounds along with a comparison between the expected number and actual number of school brackets to choose that particular combination. Note that these values reflect the expected and actual number of brackets picking the seed combination in each round, regardless of which specific No. 1, 2, 3, 4, or 9 seeded team advanced. The probability computed from the geometric distribution model shows that an Elite Eight combination of  $\{1,2,2,3,3,4,4,9\}$  seeds is not likely to occur often. This is also the case for the smaller set of four seeds  $\{1,4,4,9\}$  appearing in the Final Four.

Consequently, the expected number of the 143 school brackets submitted to have chosen these combinations is practically zero – exactly what was observed. For the finalists and eventual champion, the probability of these seeds appearing is significant enough to expect several people to have correctly made these predictions. Remarkably, the model does provide a good indication of how the student body will form predictions with reasonable accuracy.

Table 2: Expected vs. Actual Number of School Brackets with Correct Seed Combination

Round	Seed Combination	Expected Frequency (years)	Probability	Expected No. of Brackets	Actual No. of School Brackets
Elite Eight	{1,2,2,3,3,4,4,9}	3006	0.000333	0.048	0
Final Four	{1,4,4,9}	3159	0.000317	0.045	0
Finalists	{1,4}	14	0.0732	10.5	5
Champion	{1}	2	0.509	72.8	78

Since the sample size of people participating in a local school tournament challenge is rather small to adequately justify any claims about behavioral prediction, a larger, more diverse sample size would be beneficial. Fortunately, ESPN hosts a tournament challenge annually where several million brackets are submitted from fans across the U.S. For the 2013 NCAA tournament challenge, 8.15 million brackets were submitted<sup>6</sup>. Of these, only 47 correctly chose Louisville (1 seed), Michigan (4 seed), Syracuse (4 seed), and Wichita State (9 seed) to appear in the Final Four. The odds against {1,4,4,9} seeds appearing in the Final Four is 3159 to 1 according to the geometric distribution model. Since there are 12 combinations of {1,4,4,9} seeds, the odds against these four particular teams making the final four is  $12 \times 3159 = 37,908$  to 1, assuming all combination of teams are equally likely. This yields an expected number of  $8,150,000 / 37,908 = 215$  brackets to correctly choose the Final Four participants. Clearly, the population would not choose all combinations of {1,4,4,9} teams equally, as a team's fan base plays a significant factor. However, considering the magnitude of brackets submitted, the expected and actual number on the same order. Table 3 further supports this phenomenon by comparing the number of brackets who chose each Final Four representative to become champion.

Table 3: ESPN Bracket Analysis

Team	Percent Who Picked to Win	Actual No. of Correct Brackets	Probability	Expected Number of Correct Brackets
(1) Louisville	21.9%	1,784,850	0.127	1,037,088
(4) Michigan	2.7%	220,050	0.0151	122,658
(4) Syracuse	1.5%	122,250	0.0151	122,658
(9) Wichita St	0.03%	2,776	0.00043	3,505

#### IV. Engineering Implications and Future Direction

This paper presents an analysis of the NCAA basketball tournament which is entirely mathematical in its methodology. However, the modeling process and its applications have strong engineering implications. This section provides a few example topics for classroom instruction that could potentially follow this tournament analysis within a high school engineering course, along with future extensions planned for the current curriculum.

Examples of well-established engineering applications involving the use of Bernoulli trials and binomial and geometric distributions reside in the practice of quality control. In a large batch of manufactured products, it is common practice to take a small test sample of these products to assure quality requirements are met. Each sample is deemed to either pass or fail a specific test criterion. Then, the number of samples failing this test has been shown to follow the binomial distribution. If samples are periodically taken from a continuous process, then the number of samples tested until the first failure is found is geometrically distributed.

In digital communications, data is sent electronically in sequences of bits. In a perfect world, information would be transmitted with zero errors. Instead, engineers design error checking algorithms to capture and repair as many of these errors as possible. Improving the likelihood of an error occurring or passing undetected is critical in designing reliable communication systems.

Mathematical statistical modeling goes well beyond the realm of sports analysis. Numerous applications exist in engineering practice which requires the designer to account for factors of uncertainty. For example, just-in-time is an industrial engineering business strategy whose goal is to maintain an inventory of available components or products at an optimal, cost-efficient level, where production or consumer demand may fluctuate randomly but predictably. Adaptive control accounts for external environmental factors to help robotic devices decipher unknown operating parameters (e.g., airplane stabilization subject to varying cross winds). The efficiency of transportation systems are also subject to variations in traffic volume. Here, mathematical models can help determine the best plan for shipping rail cargo, the effect of road traffic light sequencing, and the determination of airline scheduling. Each of these examples must take into account factors relying on the outcomes of random events.

A time consuming aspect of implementing the school wide tournament challenge is the data entry process for all the submitted brackets. Rather than spending hours to enter over a hundred bracket selections by hand into the Excel computational spreadsheet, the creation of an automated entry system would be highly beneficial in eliminating this mundane task. The school's engineering class has recently begun teaching HTML, CSS, and JavaScript as part of its curriculum, and these newfound skills can be utilized by creating a web application where students can create and submit their bracket entries. In addition to automating the data entry process, this application can compute statistics pertaining to the school's seed selections, and provide numerical and graphical comparisons to what is predicted. Eventually, the plan also includes placing the instructional material as part of the online content. This future extension teaches the students how to generate a product built on an application of mathematical theory.

Currently, the probability values in the geometric distribution are computed using information from the past 29 tournaments (those containing a field of 64+ teams). As the game evolves both in rules and strategies, it would be interesting to understand if parity arises amongst the seeded teams. An analysis of a change in the probability that a particular seed combination appears in each round could be performed by considering a rolling window of past tournaments (i.e., computed using the past 5 or 10 tournament results). Any distinct changes could indicate the influence of noteworthy changes made by the NCAA (e.g., the introduction of the 3-point arc, equity in scholarships).

Lastly, Monte Carlo techniques can be used to generate a distribution of scoring point totals based on a bracket designed to select the most probable seed winners in each round. There are four sets of No. 1-16 seeded teams starting the tournament, with one set in each region. Since the most probable champion is a No. 1 seed according to the model, there exist four scenarios of teams in which this could occur. Similarly, the most likely finalists to compete for the championship are a No. 1 vs. No. 2 seed, generating 16 possible team matchups. If this process is continued throughout the bracket, Monte Carlo simulation can be used to investigate the distribution of total points scored based on all combinations of teams that fit the most likely seed combination in each round. This distribution of point totals can then be compared to that resulting from the general population. Such an exercise can help students understand how basing decisions on mathematical analysis can (or cannot) improve upon those made purely by intuition.

## V. Conclusions

This paper outlined the curriculum for a week long analysis of the NCAA men's basketball tournament. The instructional material is intended to supplement standard curricula taught in high school mathematics and engineering courses. The real life application of probability and statistics to sports analysis serves to generate excitement in learning and applying basic concepts, and to motivate students to explore the use of mathematics to gain a better understanding of events inherent with uncertainty. A school wide activity is discussed to show how this educational effort can be expanded outside the classroom to involve the entire student body, in the hopes of motivating students to enroll in elective courses in the future. The classroom students can then analyze the data obtained from this school wide challenge to determine if mathematical models can be used to help understand human intuition. Ultimately, this week long experience helps students realize the practical applications of mathematics, and demonstrates that a systematic analysis in lieu of intuition can give your bracket the statistical edge.

## References

1. Jacobson, S. H., Nikolaev, A. G., King, D.M., Lee, A. J., 2011, "Seed distributions for the NCAA Men's Basketball Tournament", OMEGA, 39(6):719-724.
2. Lunardi, J, 2013, "Fundamentals of Bracketology," St. Joseph University College of Professional and Liberal Studies, Online Course.
3. Ashton, J, 2013, "Lindner business analytics professor and alum bring March Madness fever to the classroom," University of Cincinnati News Release, March 7, 2013. Website available at <http://www.uc.edu/news/NR.aspx?id=17449>. Accessed Nov. 11, 2013.
4. Shishebor, Z. and Towhidi, M., 2004, "On the generalization of negative binomial distribution", Statistics and Probability Letters, 66:127-133.
5. ESPN Tournament Challenge. Website available at <http://games.espn.go.com/tournament-challenge-bracket/en/frontpage>. Accessed Nov. 11, 2013.
6. Brennan, E., 2013, "Tourney Challenge: 47 got Final Four," ESPN News Release, March 31, 2013. Website available at: [http://espn.go.com/blog/collegebasketballnation/post/\\_/id/82590/tournament-challenge-only-47-got-final-four](http://espn.go.com/blog/collegebasketballnation/post/_/id/82590/tournament-challenge-only-47-got-final-four). Accessed Nov. 11, 2013.