RANKING COLLEGE FOOTBALL TEAMS INDEPENDENT OF VICTORY

Honors Thesis

Presented in Partial Fulfillment of the Requirements
For the Degree of Bachelor of Mathematics

In the College of Arts and Sciences
at Salem State University

By
Ryan Kelly

Professor Brian Travers
Faculty Advisor
Department of Mathematics

***

Commonwealth Honors Program
Salem State University
2014

Ranking College Football Teams Independent of Victory Margins

Ryan Kelly

December 17, 2013

Abstract

This paper addresses David Mease's formula for ranking college football teams. It is just one of the numerous formulas that can be used by the Bowl Championship Series in ranking the top teams in the country. Mease uses this formula to rank teams independent of victory margins, something not all formulas take into consideration. Winning margin may help teams gain higher ranking in some formulas, so this formula ignores that statistic.

1 Introduction

This paper is about the mathematics behind the Bowl Championship Series (BCS) of college football, which is the selection system that ranks the top college football teams in the country. The system was created in 1998 after years of controversy over the rankings. The dissension occurred because people believed that the rankings were too heavily weighted on subjective opinions rather than objective statistics and facts, which led to the creation of the Bowl Championship Series. Before the establishment of the BCS, ranking college football teams consisted solely of two polls: the Associated Press poll and the Coaches poll. The new BCS ranks teams using numerous mathematical formulas based on statistics and probability in addition to the Coaches and Associated Press polls. Between the formulas the Bowl Championship Series uses and the voters that rank teams, they include many different factors when creating the rankings. One of the overarching ideas for this paper is that the BCS rankings are too heavily weighted based on *victory margins*, or how much a team won a game by.

The work done by David Mease on the mathematics behind his BCS formula is the inspiration of the paper. His formula centers around ranking college football teams independent of victory margins–a vital aspect in some ranking schemes. The model is based in statistics and probability, two major mathematical concepts for determining maximum likelihood outcomes for teams. We will look at maximum likelihood outcomes in the following sections. Although his work is important in ranking college football teams, it is just one of the mathematical formulas that are used to rank the top college football teams in the country.

The following section will illustrate an example of how to rank teams independent of victory margins and showcases arguments for this type of ranking. Then we will detail the probability theory on which Mease bases his model, followed by an in depth explanation of the formula. Following the details of the formula, we will showcase the rankings of teams today and the chances of them winning or losing games in certain situations. We will close with suggestions to improve the formula and to further ranking schemes in general. Mease's formula is one of many to examine in the Bowl Championship Series, but is the primary focus for the scope of this paper.

2 An Example

Ranking teams independent of victory margins is the inspiration behind David Mease's formula. He believed that ranking teams by ignoring a team's victory margin was the fairest way to do it.[1] A simple way of explaining why victory margins could be ignored is with the following example. Suppose there are 5 teams named A, B, C, D, and E for simplicity and they play 8 games with the following outcomes.

Games
Game 1: Team A defeated Team C.
Game 2: Team A defeated Team E.
Game 3: Team B defeated Team A.
Game 4: Team B defeated Team E.
Game 5: Team C defeated Team D.
Game 6: Team C defeated Team E.
Game 7: Team D defeated Team E.
Game 8: Team D defeated Team E.

Team	W	L
A	2	1
B	2	0
C	2	1
D	2	1
E	0	5

In the table one can clearly see each of the teams records, which consists of their wins and losses denoted W-L. Team A's record is 2-1, Team B is 2-0, Team C is 2-1, Team D is 2-1 and Team E is 0-5. Team B would attain top ranking because they are undefeated and Team E would be in last place because they went 0-5. However ranking 2nd place through 4th place is where the voters deliberate, because 3 teams finished 2-1. One could argue that Team A should be ranked highest of all the 1 loss teams because their lone loss was at the hands of the top ranked team. In similar fasion, one could reason that Team C is in third, because their lone loss is to Team A and their only loss is to Team B which leaves Team D in fourth place with their only loss being dealt by Team C. Thus the final rankings are Team B, Team A, Team C, Team D, and Team E.

Mease's reason for putting this example in his paper was to explain how complex it is to fairly rank teams solely based on only wins and losses.[1] However, it also shows the thought process behind the rankings and that it is possible to objectively rank teams without the use of margin of victory. Before the establishment of the BCS, the Coaches Poll and the Associated Press Poll voters had to rank 125 Division 1-A college football teams based on numerous categories. It was easier by comparison to rank the 5 previous teams because only 8 games were played, however it is much more difficult to rank 125 teams with between 11 and 14 games per season without the lack of common opponents and overwhelming number of games to consider.[1] This was Mease's argument to create a computer system based on a team's strength and *winning percentage*. Winning percentage is the likelihood of winning a game. So if one team played 12 games and won 9 of them, their winning percentage would be 75%.

3 Probability Model

A standard normal cumulative distribution function is a function that tells the probability that an *event* in some context will fall between any two real numbers.[5] An event is a subset of a sample space, or all possible outcomes.[2]

In general, the standard normal cumulative distribution function is given by the following equations and will be explained in the following paragraphs.

$$\Phi(y) = P(Y \leq y) = \int_{-\infty}^{y} f(x)dx$$

where

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \text{ [2], [5]}$$

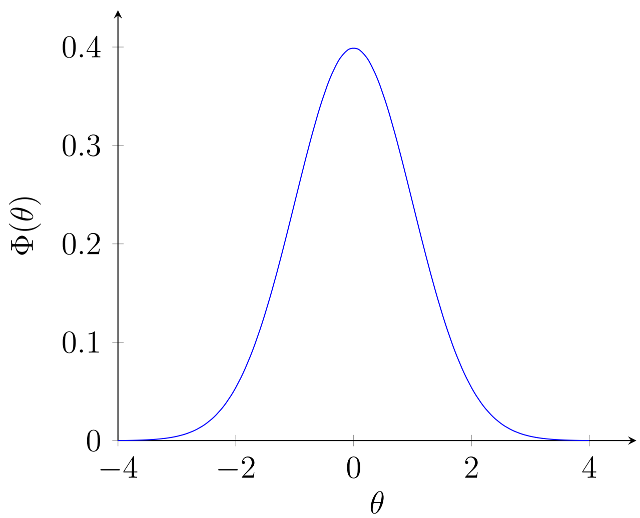
The probability model used throughout Mease's formula is a variation of the standard normal distribution function denoted Φ where Φ is the cumulative distribution and requires you to take the integral, denoted above.[5] In our case the standard normal cumulative distribution function is telling us the probability that a team will win will be between 0 and 1. That outcome is based on the domain values of the function, all real numbers between -2 and 2. [3] A standard normal cumulative distribution function has many variables. The first variable is μ and it is mean value or expected value.[2] The expected value can be considered the "center of mass" of its distribution.[6] Standard normal distributions have a mean μ value of 0. The next variable is σ and that is the variance. The variance of a function measures how far a set of numbers are spread out. [6] In standard normal distribution the σ value is 1, but this is where our variation arises, because our function's σ value is 1/2. Mease chooses 1/2 for less spread out values, because teams can be upset by lower ranked teams thus allowing a less consistent probability. [1]

The next variable is y and it is the difference between two inputs when one is trying to find the probability of their event happening. In our formula it is the difference between the *strength ratings* of teams, which is denoted by θ when finding the probability that some team will win a game. The θ is a value determined by a team's record as well as on the strength of its opponents, who in turn are judged based on their records.[1] The higher it is, the better the ranking a team has. The strength rankings have values which are all real numbers between -2 and 2 and that when substituted into the above equation yield values between [0,1], because the probability that a team beats another team can't be more than 100%. The next variable is x and is the variable by which we will be integrating. The letter

x equals $\frac{x}{\sigma}$, but since σ equals 1 in standard normal distribution it was left out in the above equation. In our case though, we know that x can be simplified to $\frac{x}{1/2}$ because σ is 1/2. The next term, $P(Y \leq y)$ states that the probability of some *random variable* being less than or equal to the difference in strength values will equal $\int_{-\infty}^{y} f(x)dx$. A random variable is a function that maps the sample space to the real line.[6] That means that Y maps y values onto the standard normal distribution curve shown below.

Taking the integral from $-\infty$ to y allows one to calculate the aforementioned winning percentage for some team. This is because the probability that a team will win is calculated by finding the area under the curve from $-\infty$ to y.

The following graph shows the standard normal distribution curve.



An example of how the integral works is the following. Let two teams play one another where Team A has a strength rating of 1 and Team B also has a strength rating of 1. One would imagine that since the two teams have the same rating they would have equally likely chances of winning. Since y is the difference in strengths, y is 0. The integral would read

$$\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x)^2} dx.$$

After solving the integral or calculating the area under the curve one does gets 50%. More in depth examples of this will be examined later in the paper.

4 The Formula

How did Mease rank teams without margin of victory? He does so using his equation, which is a *maximum likelihood function*. The likelihood of a set of parameter values, θ, given outcomes z, is equal to the probability of those observed outcomes given those parameter values. [9] This can be explained by *conditional probability*. Conditional probability is the probability that some event occurs given that another event occurs and can be denoted $P(A|B)$.[2] This reads as "the probability of event A given event B." In our case, the likelihood function uses conditional probability to see how a team is going to do, winning or losing throughout the season. We know a team's strength ranking and from that we can find out if a team wins or loses. So let z be the outcome of some game and θ be a team's strength rating. Then the likelihood of some team t winning over the course of the n game season can be denoted by

$$\Phi(z_1|\theta) \times \Phi(z_2|\theta) \times \Phi(z_3|\theta) \times \ldots \times \Phi(z_n|\theta) = \Phi(z_1, z_2, z_3, \ldots, z_n|\theta)$$

where

$$\Phi(z_1, z_2, z_3, \ldots, z_n|\theta) = \prod_{i=1}^{n} \Phi(z_i|\theta).[7]$$

A likelihood function contains all of the information about some unknown parameter.[6] Also, maximum-likelihood estimation provides estimates for the model's parameters.[7] In this particular case, it estimates how good a team is and how likely they are to win. The higher this value, the better the team. The formula has three parts in figuring out how likely a team is to win and is based off a team's strength rating.

Now we can look at Mease's formula, as follows:

$$l(\theta) = \prod_{(i,j)\in S} [\Phi(\theta_i - \theta_j)]^{n_{ij}} \times \prod_{i=1}^{n} \Phi(\theta_i)\Phi(-\theta_i) \times \Phi(\theta_{n+1})\Phi(-\theta_{n+1}) \times \prod_{(i,j)\in S^*} [\Phi(\theta_i - \theta_j)]^{n_{ij}}.$$

The formula is broken down into the following components.

$$l(\theta) = \underbrace{\prod_{(i,j)\in S} [\Phi(\theta_i - \theta_j)]^{n_{ij}}}_{Part\,1} \times \overbrace{\prod_{i=1}^{n} \Phi(\theta_i)\Phi(-\theta_i)}^{Part\,2} \times \underbrace{\Phi(\theta_{n+1})\Phi(-\theta_{n+1}) \times \prod_{(i,j)\in S^*} [\Phi(\theta_i - \theta_j)]^{n_{ij}}}_{Part\,3}$$

4.1 Part 1

First we consider Part 1 of the equation. Let S be the set of all games between teams in Division 1-A college football. There are n teams in S and let i and j denote separate teams in S. The beginning of the first part of the equation is Π. It is the product of all the likelihoods that one team beats another team throughout the season. Mease multiplies the probabilities are because the events or games played are independent.[2] However, multiplying the wins rather than adding them works well because a string of wins against different opponents gets more and more difficult as the college football season progresses and adding strengths creates for high numbers. We previously established that the probability a team beats another team can't be more 1.

The next piece of Part 1 is $\Phi(\theta_i - \theta_j)$ and it is the probability team i beats team j. The probability that team i beats a team j is calculated by using Φ which is the standard normal cumulative distribution function. The next variable, n_{ij} is the number of times that team i plays a team j. The $(\theta_i - \theta_j)$ is the difference in strengths of the two teams. The probability of team i beating team j, or likelihood of winning, is denoted by $[\Phi(\theta_i - \theta_j)]^{n_{ij}}$. [1]

Now why didn't Mease use just Part 1 of the equation? Part 1 of the equation grades teams based on wins and losses, but doesn't restrict these grades for teams. This is because teams always want the best scores hence the name *maximum* likelihood function. But that means that the strength ranking for some teams can be infinite if they don't have any losses, because their θ ranking can't be lowered. Infinite strength rankings are impossible because if would not be possible to find the probability of an undefeated team playing against another team. This is unrealistic, which leads to Part 2 of the equation.

4.2 Part 2

Part 2 *penalizes* the strength rating. That eliminates the problem of the strength rating being infinite. Part 2 as listed above is $\prod_{i=1}^{n} \Phi(\theta_i)\Phi(-\theta_i)$.

One way to understand the idea of the penalized term is by *Bayesian posterior distribution.* The Bayesian posterior distribution of a random event, or game played in this case, is the conditional probability that is assigned after the relevant evidence is taken into account.[4] The evidence in this case is the θ ranking of a team playing. The use of Bayesian distribution is meant to attribute uncertainty rather than randomness to the uncertain quantity[8] because no team is ever guaranteed a win. Part 1 is penalized by Part 2 because it "is an estimator that minimizes the posterior expected value of the function."[10] Bayesian posterior distribution contrasts with the likelihood function[4], which is the probability of the evidence given by the aforementioned $\Phi(z_i|\theta)$. The Bayesian posterior distribution is denoted $\Phi(\theta|z_i)$.[4] This is saying that the θ ranking is based on the outcome of the games played rather than what it was saying before, that the outcome of the game was based on the strength rating. The θ ranking depending on the game played penalizes the rating regardless of the outcome, because losing to a lower ranked team has negative effects to a teams strength rating and beating a lower ranked team doesn't improve a team's ranking.

The Π in Part 2 multiplies the outcomes of all the games played. Part 2 includes a team's strength rating when penalizing the score and includes the negative term to do so.

4.3 Part 3

The last part of the equation is Part 3, which is similar in many ways to Part 1, but with different teams. The first parts of the formula produces a fairly good ranking and likelihood of winning but Part 3 concludes it. The set S* consists of games between teams that are both not in Division 1-A, meaning that one team is in Division 1-A and one is not. Games between Division 1-A teams and non-Division 1-A teams are not common but do still occur and Part 3 covers these games. Non-Division 1-A teams are at a different level due to many factors, including size of the school, revenue, number of scholarships, and other factors. These teams, therefore can not compete at the same level as Division 1-A teams, yielding much lower θ values. However sometimes upsets occur, which is why this part is included in the formula. Part 3 of the equation as listed above is $\Phi(\theta_{n+1})\Phi(-\theta_{n+1}) \times \prod_{(i,j)\in S^*}[\Phi(\theta_i - \theta_j)]^{n_{ij}}$. Similar to Part 1, $[\Phi(\theta_i - \theta_j)]^{n_{ij}}$ has the same meaning based on Φ, which again is a standard normal cumulative distribution function. The $\Phi(\theta_i - \theta_j)$ is still the probability that team i beats team j and n_{ij} is again the number of times the teams play. The $\prod$ in front again multiplies all the games in which a team in Division 1-A plays a non-Division 1-A team.

Now to understand how they are different we must let team i be the team in Division 1-A, where they are one of the n aforementioned teams of Division 1-A and can be represented in the set $\{1, ..., n\}$. Now we can let team j be the non-Division 1-A team and can be represented by $n + 1$, because team j is not in Division 1-A thus it is not a part of the n teams. A team in Division 1-A losing to a team in Division 1-A is not a favorable outcome, but it is not as hindering as a loss to a non-Division 1-A team. A Division 1-A team's θ ranking won't increase significantly with a win over a non-Division 1-A, however a loss to that team will have a very negative effect on the team's θ ranking. This is true because of the penalizing factor, $\Phi(\theta_{n+1})\Phi(-\theta_{n+1})$. This factor, just like the way part 2 penalized part 1, alters a team's θ ranking for better or worse. A win for the Division 1-A team i takes into account the θ ranking for the non-Division 1-A team j as denoted by (θ_{n+1}), which will offer little θ value to team i, however losing to the team which a much lower θ ranking will lower the Division 1-A teams significantly denoted by the non-Division 1-A team j's θ ranking, $(-\theta_{n+1})$.

5 Teams Today and Their Rankings

Mease uses his formula to rank all 125 teams in Division 1-A college football and lists the rankings on his website, which he updates every week of the season.[3] This section is devoted to explaining the concept of standard normal distribution and demonstrates the significance of the probability with some various examples. Now as previously mentioned, the standard normal cumulative distribution function is denoted by

$$\Phi(y) = P(Y \leq y) = \int_{-\infty}^{y} f(x)dx$$

where

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

The probability that one team i beats another team j as previously stated is denoted $[\Phi(\theta_i - \theta_j)]^{n_{ij}}$. Because all of the following examples will involve teams playing once, let n_{ij}=1. We also stated previously that y is the difference between the strengths of two teams denoted $\theta_i - \theta_j$ in the formula and σ=1/2.

We can now find out some team's probability of winning a game. Closely ranked teams will have almost equally likely chances of winning and teams ranked far apart won't have the same equally likely chances. Let us examine some match ups. These rankings are based on the November 30th, 2013 Bowl Championship Series rankings.[3] Looking at the number one and two teams in the BCS, the highest two ranked teams in the country, we have the Ohio State Buckeyes and Florida State Seminoles respectively. Ohio State's strength rating, θ is 1.5467 and the θ value for Florida State is 1.5250. We will denote Ohio State's strength rating θ_i and Florida State's θ_j. Now in this particular case y is $1.5467 - 1.5250 = .0217$. So substituting our values gives us the the following integral.

$$\int_{-\infty}^{.0217} \frac{1}{(1/2)\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x}{(1/2)})^2} dx.$$

By solving the integral, the equation yields .504328, which means Ohio State has about a 50.4% chance of winning.

Let us look at another opponent for Ohio State, this time a much lower ranked team, the Georgia State Panthers. This game is between the best and the worst team in college football. Again, Ohio State's θ_i ranking is 1.5467 and Georgia State's θ_j ranking is -1.9127. Substituting our y value of $1.5467 - (-1.9127) = 3.4594$ we get the following integral.

$$\int_{-\infty}^{3.4594} \frac{1}{(1/2)\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x}{(1/2)})^2} dx.$$

After solving the integral, it yields $\approx$ .958158, which means Ohio State has about a 95.8% chance of winning. Because Georgia State has a 4.2% chance of winning, there is very little chance that Ohio State loses the game and the only reason they have a chance is because we chose σ=1/2 for the variance to allow the "upset."

There is one more game of importance to look at. It takes place between the 8th ranked team in the country, the Stanford Cardinals and the 55th ranked team in the country, the Utah Utes. The 8th ranked team in the country, Stanford, was upset by Utah this season. This loss ended Stanford's hopes of playing for the national title. Stanford's θ_i ranking is 1.1697 and Utah's θ_j ranking is .1676. Substituting our y value of $1.1697 - .1676 = 1.0021$ we get the following integral.

$$\int_{-\infty}^{1.0021} \frac{1}{(1/2)\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x}{(1/2)})^2} dx.$$

Solving the integral yields $\approx$ .691832, which means Standford had about a 69.1% chance of winning. Stanford was the favorite, they were ranked higher and had only a 30.9% chance of losing to Utah.

6 Possible Modifications

Although the formula works in estimating the likelihood of teams winning games, it is only based on teams winning or losing independent of victory margins. It successfully produces consistent rankings similar to the Associated Press and Coaches Polls but there are many other categories that could be included into the formula. These include whether a team played home or away, when the game was played, and even incorporating margin of victory. Including where a team played is a simple part to change in the equation.

By offering some tiny numerical value to a variable say, ω, and adding or subtracting it when one is finding out the difference in strengths can be an easy way to change the equation. College football games draw huge crowds and the fan support can be overwhelming for some teams. As previously stated we know that the probability that some team i beats some team j is denoted $\Phi(\theta_i - \theta_j)$. Now if we want to include some team's home field advantage we can incorporate the ω value. If team i was the home team, the probability that team i would win could be be found by $\Phi(\theta_i - \theta_j + \omega)$. If team i was the away team, the probability that team i would win could be be found by $\Phi(\theta_i - \theta_j - \omega)$.[1] It is a small fix but adds an important real life factor to the equation.

The next possible modification could be including when the game was played. Games played later in the season are more important than the games played earlier, because teams have less room for error if they want to hold on to their rankings or raise their rankings. In Mease's equation he has a term n_{ij}, which again means the number of times that team i played team j. Teams usually only play each other once per season which is why n_{ij}=1 throughout the season and the aforementioned examples. However, we could assign a higher value for n_{ij} if we wanted to make it seem more important than an earlier season game. For the same reason, we could assign a lower value for n_{ij} if we wanted to make it seem less important than an earlier season game.[1] The reason for this is that a highly ranked team may have a win against a lesser ranked opponent late in the season.

The last modification may seem a bit of a stretch but is something that could be considered. Although Mease would have to change his entire formula and paper for that matter, victory margins are vital in other formulas and adding that to his could even better help show how good or bad a team is. He has shown that his formula did not need one to create accurate rankings, but adding a small scoring differential variable, so a number that stood for how many points a team scored minus how many they gave up, into the already existing formula could give an even more accurate ranking. There is no such thing as a perfect formula in the BCS because all the formulas together make up the ranking system. However one that produces accurate rankings without margin of victory involved could almost be perfected if there was an aspect in the formula that did include margin of victory.

7 Future Work

It may be difficult to completely digest Mease's formula, but it provides insight into the way the BCS rankings are calculated. While a lot of time has been set aside for this project, much more time will be needed to finish and comprehend all of the math behind the other formulas.

David Mease's formula was just one of numerous mathematical aspects of the BCS and how they rank the top college football teams. There is much more that can be done regarding this topic and subject. The ranking of teams independent of victory margins, although key to the ranking system is just one of the numerous factors that are taken into account. Future work in this field is something that would not be hard to find. Analyzing the Coaches Poll and the Associated Press Poll and finding out how heavily weighted they are is a future work topic to consider. A preliminary question of interest and something to consider someday was "Are the BCS rankings too heavily based on AP and Coaches poll rankings and not heavy enough on the objective statistics to make the rankings unbiased?" But other than that, there are many other formulas used in the Bowl Championship Series formula that could offer further insight on this topic.

Another question that seems interesting is "Is it possible to rank different teams from different seasons fairly?" It is easy to take rankings of different teams over the 15 years the Bowl Championship Series has existed and see which teams had the higher rankings but is that fair? A team from 2001 won't be able to play a team from 2013, but can one fairly estimate which was better? Maybe someday we will be able to build a technology that could successfully simulate games between the best teams of all time and play to see who was the greatest team of all time. Being a big Notre Dame fan, it would be intriguing to see their greatest team of all time versus the greatest college football team a great school like Alabama had to offer. Unfortunately we are years away from developing such technology.

References

[1] D. Mease, A Penalized Maximum Likelihood Approach for the Ranking of College Football Teams, *American Statistician* **57 (4)** (2003), 241-248.

[2] G. Grimmett, D. Stirzaker, Probability and Random Processes, *Oxford* **Third Edition** (2001), Various pages.

[3] D. Mease, David Mease's Football Rankings, *Web, http://www.davemease.com/football/* (2013).

[4] Posterior probability, *Web,http://en.wikipedia.org/wiki/Posterior_probability_distribution* (2013).

[5] Normal distribution, *Web, http://en.wikipedia.org/wiki/Normal_distribution* (2013).

[6] K. Knight, Mathematical Statistics, *Chapman & Hall/CRC* **First Edition** (2000), Various pages.

[7] Maximum likelihood, *Web, http://en.wikipedia.org/wiki/Maximum_likelihood* (2013).

[8] Prior probability, *Web, http://en.wikipedia.org/wiki/Prior_probability* (2013).

[9] Likelihood function, *Web, http://en.wikipedia.org/wiki/Likelihood_function* (2013).

[10] Bayes estimator, *Web, http://en.wikipedia.org/wiki/Bayes_estimator* (2013).