2010-2011

Project Number: JPA 1004

Worcester Polytechnic Institute

Danielle Beland
Weixin Chen
Brannon Cote-Dumphy

Project Advisor: Jon Abraham

A Major Qualifying Project Report

submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

[ANALYSIS AND IMPROVEMENT OF FANTASY FOOTBALL TEAM MANAGEMENT SOFTWARE]

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Abstract

Our project was sponsored by Advanced Sports Logic to assist in the development of fantasy football team management software. We helped develop mathematical formulas and statistical models that will allow for the most precise approximations given the constraints imposed by the computer on which the software is implemented, as well as the time limit imposed in a fantasy football draft. The software is designed to assist fantasy football team owners in selecting players in a draft, as well as provide trade, keeper, waiver, and starting lineup recommendations throughout the season, to give them the best chances of winning their league.

Table of Contents

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Table of Figures

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Executive Summary

Advanced Sports Logic is an entrepreneurial company which provides its customers with software to guide the user in improving their chance to win a fantasy football league. The goal of the software is to apply rigorous, mathematically correct methods to provide recommendations for all transactions possible within a typical fantasy football league. The project was undertaken to both verify the accuracy of these methods and to develop newer, more refined approaches where possible.

The goal of this project was to both provide Advanced Sports Logic Inc. with an approach to improve their current methods, as well as describe their existing process in common mathematical terms. Steps included:

- Meeting with the CEO / Founder to discuss the current software and formulas
- Creating a new diagram of the process flow within the software
- Writing summaries of the current formulas and techniques in understandable mathematical terms
- Writing descriptions of mathematically correct methods to determine certain data
- Suggesting recommendations to improve the current system

Success regarding this project required a full understanding of the current software. Once a comprehensive analysis of the software was complete, Advanced Sports Logic Inc. requested a flow diagram of the current methods. The flow diagram outlined the major phases of the software and was accompanied by written descriptions which elaborated upon the more sensitive areas within the process. These descriptions included analyses of the current mathematical formulas. Using our knowledge of probability and statistics, we analyzed the current methods with the utmost scrutiny to determine their mathematical accuracy. Although most elements of the software were mathematically sound or used reasonable approximate methods, we were able to identify some areas for improved mathematical rigor. We then sought ways to improve these methods, within the limits of available computational power, and provided Advanced Sports Logic Inc. with recommendations for further increasing the accuracy and understandability of their software.

Advanced Sports Logic Inc. can use the results of this project to help them acquire a patent for their software, and can use the data obtained to test their current methods against the mathematically-precise methods developed by us in order to determine the software's accuracy. The project provides Advanced Sports Logic Inc. with an opportunity to better understand and improve their product.

Introduction

Fantasy football is an interactive, virtual competition in which individuals act as team managers, draft professional football players and face other fantasy teams within their league. In addition to drafting players, team managers also make trades, add or drop players and change rosters. It is estimated that there are over 18 million individuals who play fantasy football. There are many different types of leagues, and each league has its own set of rules. For example, in some leagues the winner is determined based on weekly head-to-head matchups with other fantasy football teams, while in other leagues the winner is based on the total number of fantasy points his team scores throughout the season. There are many websites which host fantasy football leagues, including NFL.com, Yahoo! Sports, CBSSports, and myfantasyleague.com. For more information, please refer to Appendix A.

Advanced Sports Logic was recently formed with the goal of assisting fantasy football team managers using mathematical analysis. It provides its customers with software, known as “The Machine”, which the customer can use to improve their chance of winning their fantasy football league. “The Machine” uses a system of processes in the “Season Calculator”, which we have broken into eight phases (See Figure 1). Within those eight phases, the season Calculator uses ten functions, which have been named simply “Function 1” through “Function 10”.

The purpose of our project was to analyze the methods and functions that were in place, and describe them in common, mathematical terms. Through the course of our project, we worked to identify what would be the most mathematically accurate way to calculate, for example, what the value of a backup player would be or how to determine what seed a team might be ranked. Some of these methods would require very large numbers of calculations, making them impractical for the software. For example, calculating a team’s probability of being a certain seed, which then makes it to the playoffs, in a twelve team league, would require thousands of trillions of calculations. Since the season calculator needs to be able to run during a draft and aid in the selection of players, our methods needed to be modified, to be usable given the constraints of both the time limit during the draft, as well as the constraints imposed by the client’s computer on which the calculations are performed.

Recommendations for future work are contained later in this report.

Description of Software

The processes taken by the season calculator can be broken down into the following phases:

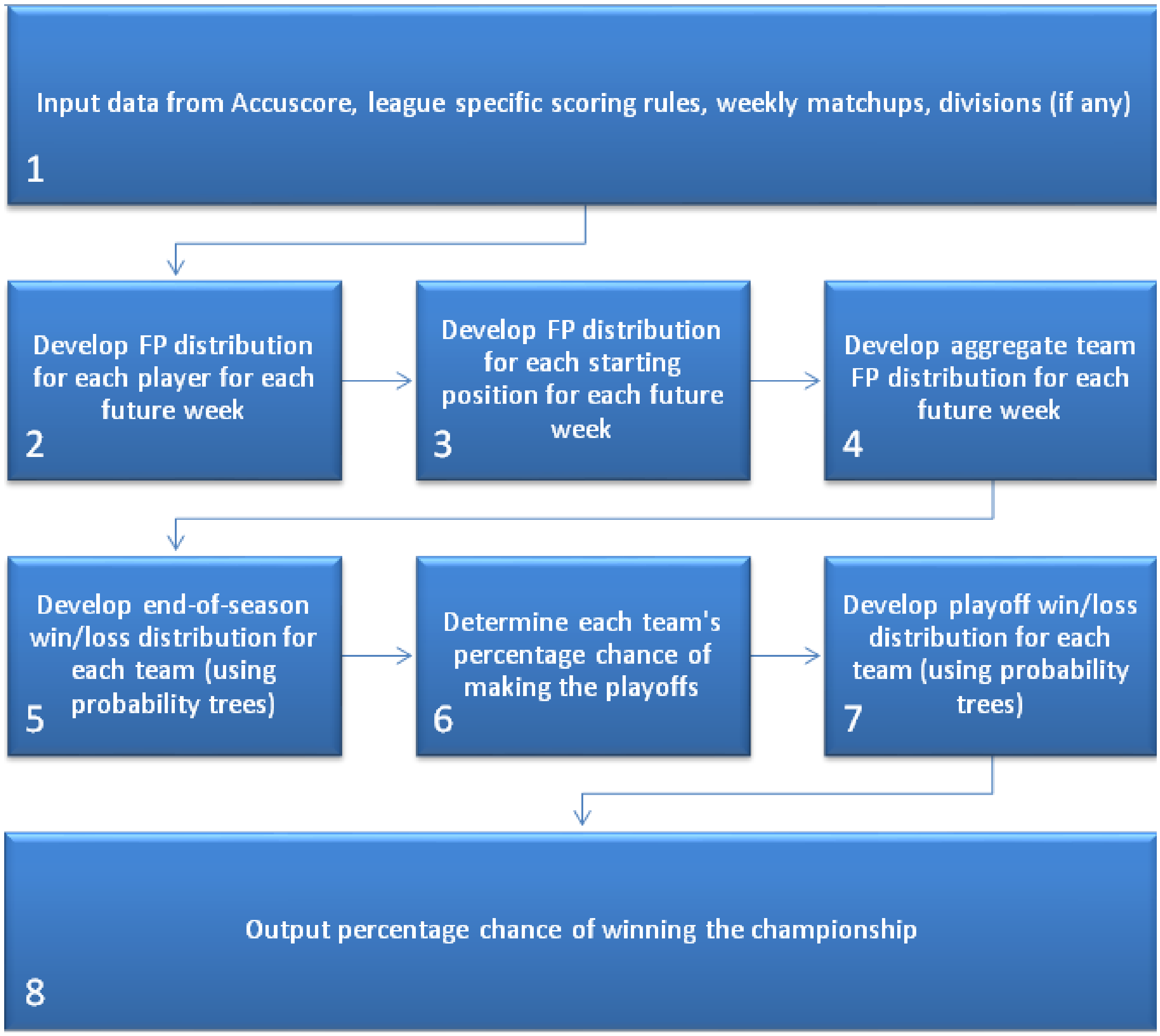


Figure 1: Phases of Season Calculator

Phase 1: Input Data from AccuScore, League Specific Scoring Rules, Weekly Matchups, Divisions (if any)

Overview:

Before the Season Calculator can begin any processes, and in order to use any of the functions from Function 1 through Function 10, some preliminary steps must first be performed:

I. Calculate the fantasy point normalization factor
II. Calculate weekly player projections based on league-specific scoring rules
III. Calculate the league-specific 1-week STDpp (a measure of the variance associated with the player projections; as time goes on, the projections will change)
IV. Calculate the weekly STDpp values
V. Find the max bin value for each position (most points scored by a player at that position in any single week)
VI. Enter league-specific data (teams, matchups, divisions)

I. Calculate the fantasy point normalization factor

Purpose:

The purpose of the fantasy point normalization factor is to convert fantasy point projections from AccuScore's scoring system to the league's scoring system.

Requirements:

I. Weekly fantasy point projections using AccuScore's scoring rules
II. Full season stats
III. League scoring rules

Process:

AccuScore provides weekly fantasy point projections for the whole season, on a player by player basis. For more information about AccuScore, please see Appendix B. For example, for all NFL quarterbacks, AccuScore provides:

Projected pass completions
Projected pass attempts
Projected percent passes complete
Projected passing yards
Projected passing touchdowns
Projected interceptions
Projected times to be sacked
Projected rushing yards

Using league-specific scoring rules and full season stats provided by AccuScore, one can calculate the total fantasy points a player is projected to score for the entire season. By calculating the total fantasy points for the league and then dividing by the total fantasy points projected by AccuScore, a normalization factor is created.

$$\text{Normalization factor} = \frac{total\ fantasy\ points\ based\ on\ the\ league\ scoring\ rules}{total\ fantasy\ points\ based\ on\ the\ AccuScore\ scoring\ rules}$$

$$= \frac{sum\ of\ all\ fantasy\ points\ for\ each\ category\ for\ the\ league}{total\ fantasy\ points\ based\ on\ the\ AccuScore\ scoring\ rules}$$

II. Calculate weekly player projections based on league-specific scoring rules

Purpose:

The purpose is to get weekly projections for each player based on league-specific scoring rules.

Requirements:

I. Weekly fantasy point projections from AccuScore
II. Normalization factor

Process:

Multiply each of the player's weekly projections by their fantasy point normalization factor to yield league-specific fantasy point projections for each player.

III. Calculate the league-specific 1-week STDpp

Purpose:

The purpose is to get a standard deviation of the change of the projections. The average standard deviation of projection changes in a 1 week period is denoted as STDpp.

Requirements:

I. Weekly fantasy point projections based on league-specific scoring rules

Process:

Refer to: How to calculate STDpp and STDpa

Example:

Refer to: How to calculate STDpp and STDpa

II. Calculate the weekly STDpp values

Purpose:

The purpose is to get a standard deviation of the change of the projections for all weeks, not just one week. Denote this as a one-sigma value.

Requirements:

I. 1-week STDpp

Process:

Multiply STDpp by the square root of the number of weeks out. For example, in week 4, the standard deviation of a specific position for week 6 is:

$$one - sigma\ value = \sqrt{(6-4)} * STD_{pp}$$

III. Find the max bin value for each position

Purpose:

The purpose is to get a maximum bin value for a specific position.

Requirements:

I. Maximum expected fantasy point projection for each position, denoted as M
II. STDpa (a measure of the accuracy associated with player projections)
III. STDpp (a measure of the variation associated with player projections; as time goes on, the projections will change)

Process:

$$Max\ Bin\ Value = M + max\left(2 * STD_{pa}, \sqrt{remaining\ weeks} * 2 * STD_{pp}\right)$$

IV. Enter league-specific data (teams, matchups, divisions)

Purpose:

The purpose is to give the Season Calculator an accurate description of the user's league to allow for maximum accuracy.

Requirements:

I. List of other teams in the league and their updated player rosters (post-draft)
II. Matrix of all matchups taking place within the league
III. Organization of teams by division (if divisions exist within the league)

Process:

Advanced Sports Logic Inc. is currently working to form relationships with a few major fantasy football providers so that the league-specific information does not need to be entered by the user. If the user chooses to utilize a different provider, they must manually enter all of the league-specific data into the Season Calculator. This data includes (but is not limited to): team-specific player rosters, divisions, weekly matchups for all teams in the league, starting rosters, and trades as they occur.

Phase 2: Generate a Probability Distribution Based on Projections for Each Player for Each Week

Purpose:

The purpose is to generate fantasy point distributions for each player for all future weeks.

Requirements:

I. DT2 (fantasy points bin values)
II. DT1 (probabilities corresponding to bins)
III. STDpa (for the current week)
IV. STDpp (for all other weeks)

Process:

The process of generating projections probability distribution for each player for each week requires the use of Function 1. It requires both Data Types 1 and 2, which are defined as follows:

Data Type 1 (DT1), Data Type 2 (DT2)

Definition:

Data Type 1 is related to probabilities. It is the probability associated with each bin of fantasy points.

Data Type 2 is the fantasy point bin value. It is the range in which each probability will fall on the horizontal axis. In this software 16 bins are used, and bins are always evenly spaced.

Example:

Here is an example of using a normal distribution with mean 22.5 and a standard deviation of 7. The third column, labeled F(X≤Mid Point) is the cumulative distribution function of a normal distribution, while the fifth column, labeled Probability, represents the difference between adjacent rows of the third column (so that it represents the probability of an outcome occurring within each bin).

Point	Mid Point	F(X≤Mid Point)	**Bin Value**	**Probability**	mean	22.5
0	1.5	0%	**(-∞,3]**	**0%**	σ	7
3	4.5	1%	**(3,6]**	**0%**		
6	7.5	2%	**(6,9]**	**1%**		
9	10.5	4%	**(9,12]**	**3%**		
12	13.5	10%	**(12,15]**	**6%**		
15	16.5	20%	**(15,18]**	**10%**		
18	19.5	33%	**(18,21]**	**14%**		
21	22.5	50%	**(21,24]**	**17%**		
24	25.5	67%	**(24,27]**	**17%**		
27	28.5	80%	**(27,30]**	**14%**		
30	31.5	90%	**(30,33]**	**10%**		
33	34.5	96%	**(33,36]**	**6%**		
36	37.5	98%	**(36,39]**	**3%**		
39	40.5	99%	**(39,42]**	**1%**		
42	43.5	100%	**(42,45]**	**0%**		
45	46.5	100%	**(45,∞)**	**0%**		

Figure 2: An Example of the Normal Distribution

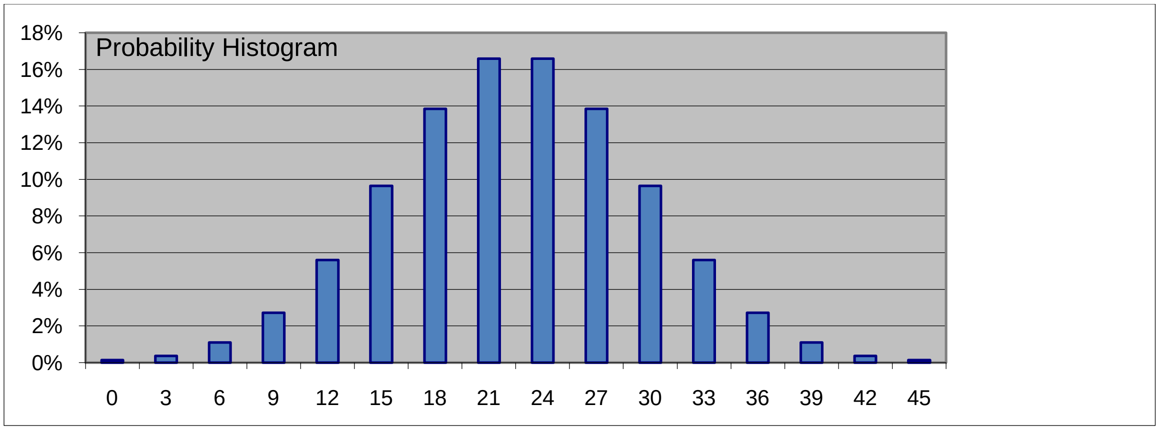


Figure 3: Probability Histogram of the Normal Distribution

Phase 3: Develop a Fantasy Point Distribution for Each Starting Position for Each Future Week

Purpose:

In order to find the value of a certain position within a fantasy football team, it is first necessary to generate the value of having backup players on your roster for that position. Depending on the number of starters for that particular position this could turn out to be computationally intensive, and it is possible that our proposed method may be impractical given the current method of implementation for the "Season Calculator".

Process:

Suppose you have two players who can each fill a single starting position. Each player's distribution is broken up into 16 bins, to make the data more manageable. Function 3 calculates the "better of" two players by finding the maximum value of all combinations of the two player projections, with associated probabilities, to yield the distribution of the "better of" curve.

Currently, Advanced Sports Logic uses Function 3, as well as the Function 3i, to identify both the "best of" and "worst of" players respectively. This scheme is run recursively to create the best combination of players.

Proposed Method:

Suppose in the running back case, we have three players, two of them are starters, and the other one is a back-up. Here are their distributions (note that Player B's distribution is quite unrealistic in the context of fantasy football projections, but is illustrated this way to make the example a little more interesting):

	A	B	C
1	0.01389	-	-
2	0.02778	-	0.03000
3	0.04167	-	0.05000
4	0.05556	-	0.06000
5	0.06944	0.10000	0.07000
6	0.08333	0.20000	0.08000
7	0.09722	0.30000	0.09000
8	0.11111	0.40000	0.10000
9	0.11111	-	0.10000
10	0.09722	-	0.09000
11	0.08333	-	0.08000
12	0.06944	-	0.07000
13	0.05556	-	0.06000
14	0.04167	-	0.05000
15	0.02778	-	0.03000
16	0.01389	-	0.04000

Figure 4: Players A, B, C distributions

The leftmost column represents the bins and the next three columns represent the probabilities associate with each bin. The bins can be converted into fantasy point ranges by multiplying a constant with evenly spaced bins. For example, the sixth bin corresponds to (6-1)*3=15 fantasy points if the sixteenth bin is 45 fantasy points.

The problem is to find the value of having backups to improve chances of scoring higher fantasy points with higher probabilities. Using a coin toss example, suppose we use the result of a coin toss on a betting game where you earn one dollar if a head shows up, and lose one dollar if a tail shows up. Assuming it is a fair coin then you have 50% chance of earning one dollar and 50% chance of losing one dollar. Now suppose the rules are changed where you have two coins and you are allowed to choose the "better" coin (after they have been tossed). The possible outcomes are still H and T, but the probabilities associated with each have been changed. Now we have four outcomes HH, HT, TH, TT, where only TT will forced us to lose one dollar. So the probability of earning one dollar is improved to be 75% where losing one dollar is decreased to 25% - assuming two fair coins flip independently.

The analogy of coin toss can be applied to having the backups on your team and improving your probabilities of scoring high fantasy points. Since there are three players each with 16 possible bins they can score, we have a total of 16^3=4096 possible outcomes, and with each outcome, we are allowed to choose the best and the better one in terms of fantasy points. Just for an illustration:

# of combination	Player A	Player B	Player C
1	1	1	1
2	1	1	2
3	1	1	3
4	1	1	4
5	1	1	5
6	1	1	6
7	1	1	7
8	1	1	8
9	1	1	9
10	1	1	10
11	1	1	11
12	1	1	12
13	1	1	13
14	1	1	14
15	1	1	15
16	1	1	16
17	1	2	1
18	1	2	2
19	1	2	3
:	:	:	:
4094	16	16	14
4095	16	16	15
4096	16	16	16

Figure 5: Combinations of A, B, C

Starting from the left column, the first column counts the number of outcomes, the second column is the bin where Player A is projected to score, the third column is the bin where Player B is projected to score, and the fourth column is the bin where Player C is projected to score. The list is arranged in such a way so that we can systematically exhaust all possible outcomes by the time we get to 4096th row. The table could be achieved using Excel where the:

Second column is: (# of combinations) MOD (16^3)

Third column is: (# of combinations) MOD (16^2)

Fourth column is: (# of combinations) MOD (16^1)

By the time we have a complete list of outcomes up to 4096th row, we can calculate our best and 2nd best players' fantasy points (by adding them) and calculate the probabilities associated with each one of the outcomes. For instance, the outcome of (A=8, B=7, C=6) would have a total of 8+7=15 points, and have a probability of 0.11111x0.30000x0.08000, or 0.002666. If the bin width is greater than one, the total points would be adjusted accordingly.

Once we calculate all the probabilities for each outcome, we have a complete distribution of "three choose two" (three bench players where the team manager can choose two starters). Here is what the distribution would look like:

Row Labels	Sum of PROB
2	-
3	-
4	-
5	-
6	0.000028
7	0.000236
8	0.001069
9	0.003292
10	0.007944
11	0.015389
12	0.026542
13	0.040472
14	0.058250
15	0.075528
16	0.091750
17	0.097194
18	0.096361
19	0.093028
20	0.087125
21	0.078542
22	0.066875
23	0.052125
24	0.037972
25	0.024583
26	0.017639
27	0.012083
28	0.007778
29	0.004583
30	0.002361
31	0.000972
32	0.000278

Figure 6: Distribution of Sum of Two Best

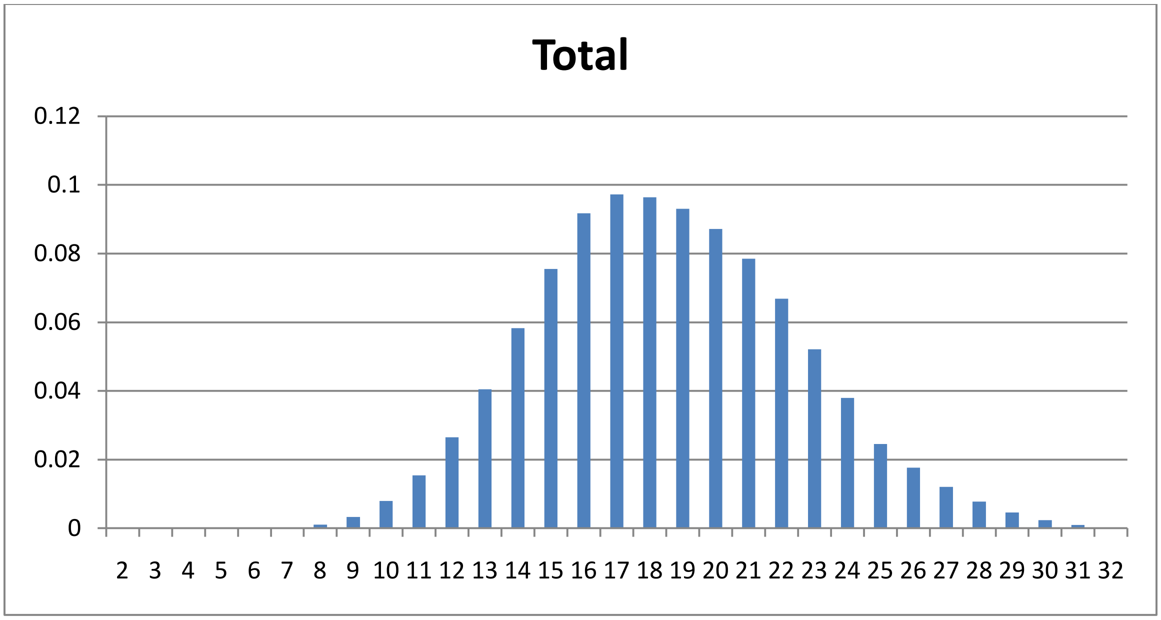


Figure 7: Graph of Distribution of Sum of two best

We could compare this distribution to the two-player distribution of only Player A and Player B. The difference is the "value" that having a back-up player (Player C in this case) adds to the team.

Here are some statistics for this example:

Statistical Analysis:				
X	X^2	P(X=x)	E(X)=sum(X*P(X=x)	E(X^2)=sum(X^2*P(X=x))
2	4	0	0	0
3	9	0	0	0
4	16	0	0	0
5	25	0	0	0
6	36	2.7778E-05	0.000166667	0.001
7	49	0.00023611	0.001652778	0.011569444
8	64	0.00106944	0.008555556	0.068444444
9	81	0.00329167	0.029625	0.266625
10	100	0.00794444	0.079444444	0.794444444
11	121	0.01538889	0.169277778	1.862055556
12	144	0.02654167	0.3185	3.822
13	169	0.04047222	0.526138889	6.839805556
14	196	0.05825	0.8155	11.417
15	225	0.07552778	1.132916667	16.99375
16	256	0.09175	1.468	23.488
17	289	0.09719444	1.652305556	28.08919444
18	324	0.09636111	1.7345	31.221
19	361	0.09302778	1.767527778	33.58302778
20	400	0.087125	1.7425	34.85
21	441	0.07854167	1.649375	34.636875
22	484	0.066875	1.47125	32.3675
23	529	0.052125	1.198875	27.574125
24	576	0.03797222	0.911333333	21.872
25	625	0.02458333	0.614583333	15.36458333
26	676	0.01763889	0.458611111	11.92388889
27	729	0.01208333	0.32625	8.80875
28	784	0.00777778	0.217777778	6.097777778
29	841	0.00458333	0.132916667	3.854583333
30	900	0.00236111	0.070833333	2.125
31	961	0.00097222	0.030138889	0.934305556
32	1024	0.00027778	0.008888889	0.284444444
Expectation	**18.5374**			
Variance	**15.5149**			
Standard deviation	**3.9389**			

Figure 8: Statistics of Distribution

Simulation:

. When the number of players (starters and back-ups) gets larger, the approach described above of enumerating every possible outcome becomes computationally difficult, and may require too much computer time to be practical. An alternate approach could be to perform a simulation. If we did this for the “three choose two” example above, we would generate a random sample of Player A’s, Player B’s, and Player C’s outcomes (based on their respective probability distributions), and then we would choose the best two out of three for that sample. Repeating this process a large number of times, we could develop an estimated distribution of the possible outcomes. This technique would allow some control on how much computer time is required, but would be somewhat less accurate than the exhaustive development described earlier.

For example: Given the distributions, we calculate their Cumulative Distribution Functions (CDFs):

	A	B	C	CDF of A	CDF of B	CDF of C
1	0.01389	-	-	0.01389	-	-
2	0.02778	-	0.03000	0.04167	-	0.03000
3	0.04167	-	0.05000	0.08333	-	0.08000
4	0.05556	-	0.06000	0.13889	-	0.14000
5	0.06944	0.10000	0.07000	0.20833	0.10000	0.21000
6	0.08333	0.20000	0.08000	0.29167	0.30000	0.29000
7	0.09722	0.30000	0.09000	0.38889	0.60000	0.38000
8	0.11111	0.40000	0.10000	0.50000	1.00000	0.48000
9	0.11111	-	0.10000	0.61111	1.00000	0.58000
10	0.09722	-	0.09000	0.70833	1.00000	0.67000
11	0.08333	-	0.08000	0.79167	1.00000	0.75000
12	0.06944	-	0.07000	0.86111	1.00000	0.82000
13	0.05556	-	0.06000	0.91667	1.00000	0.88000
14	0.04167	-	0.05000	0.95833	1.00000	0.93000
15	0.02778	-	0.03000	0.98611	1.00000	0.96000
16	0.01389	-	0.04000	1.00000	1.00000	1.00000

Figure 9: PMF and CDF of Distribution

Next, we generate a random number on (0,1) – most software programs have this functionality built in. Suppose the uniform number we generate is 0.457. Then Player A is assigned to the 8th bin because 0.38889<0.457≤0.5 (this is all based on his CDF). Then we generate another uniform number and assign Player B to a bin, and a third uniform number to assign Player C to a bin. In order to generate a single combination of Player A, B, and C's values, we require three uniform random numbers. Below is an example of 20 simulations using Excel.

	Simulation			Pr(X<=x)			Compute Best of 2		
Simulations	A	B	C	A	B	C	1st	2nd	Sum 2 best
1	0.48	0.19	0.28	5	7	3	7	5	12
2	0.27	0.10	0.88	7	7	8	8	7	15
3	0.99	0.49	0.48	6	8	13	13	8	21
4	0.16	0.86	0.94	9	7	8	9	8	17
5	0.52	0.84	0.43	3	8	7	8	7	15
6	0.65	0.14	0.33	16	7	15	16	15	31
7	0.93	0.76	0.83	11	8	9	11	9	20
8	0.56	0.25	0.01	9	7	15	15	9	24
9	0.72	0.46	0.25	8	8	11	11	8	19
10	0.47	0.90	0.14	10	8	11	11	10	21
11	0.58	0.50	0.55	6	6	16	16	6	22
12	0.46	0.14	0.98	10	8	13	13	10	23
13	0.03	0.86	0.18	8	8	10	10	8	18
14	0.97	0.69	0.54	14	5	11	14	11	25
15	0.02	0.76	0.62	4	8	11	11	8	19
16	0.46	0.22	0.67	11	8	10	11	10	21
17	0.43	0.15	0.15	8	8	6	8	8	16
18	0.59	0.25	0.27	7	8	13	13	8	21
19	0.38	0.29	0.93	3	7	9	9	7	16
20	0.79	0.32	0.07	9	6	10	10	9	19

Figure 10: Simulation of Results

Once we get up to maybe 1000 simulations, we can count how frequently the "sum of two best" occurs and will have a distribution of the result. You can compare Figure 12 below to Figure 7 (earlier) and see that the simulation approach closely approximates the "exact" approach for this situation:

OUTPUT HERE		n=	2000
Possible Values	Count Occurances		Probability
1	0		0
2	0		0
3	0		0
4	0		0
5	0		0
6	0		0
7	0		0
8	1		0.0005
9	8		0.004
10	14		0.007
11	15		0.0075
12	44		0.022
13	80		0.04
14	123		0.0615
15	143		0.0715
16	203		0.1015
17	194		0.097
18	181		0.0905
19	175		0.0875
20	183		0.0915
21	138		0.069
22	122		0.061
23	114		0.057
24	83		0.0415
25	61		0.0305
26	43		0.0215
27	34		0.017
28	15		0.0075
29	14		0.007
30	8		0.004
31	3		0.0015
32	1		0.0005

Figure 11: Simulation of Distribution of Sum of Two Best

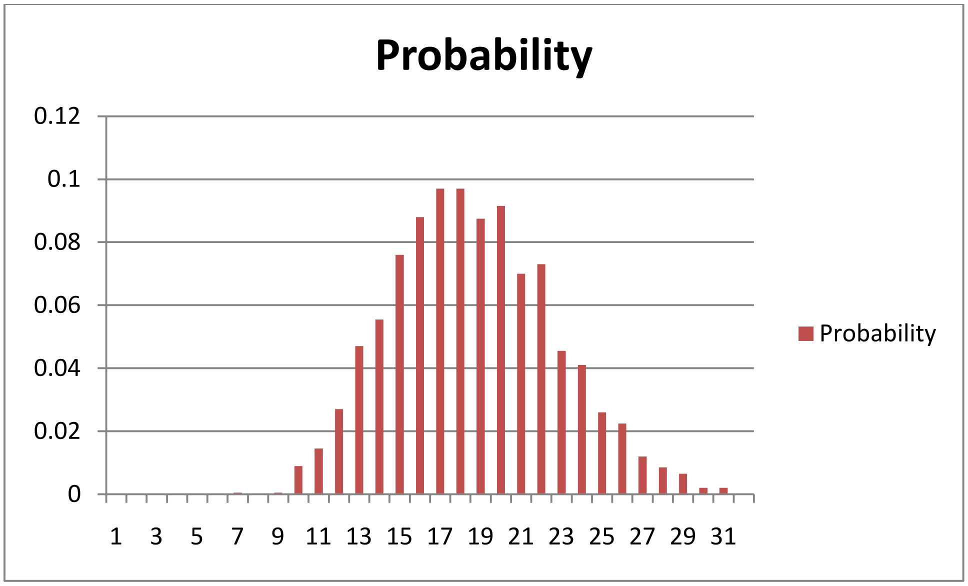


Figure 12: Graph of Simulation of Sum of Two Best

Below is a graph that shows how closely the simulation relates to the exact probabilities:

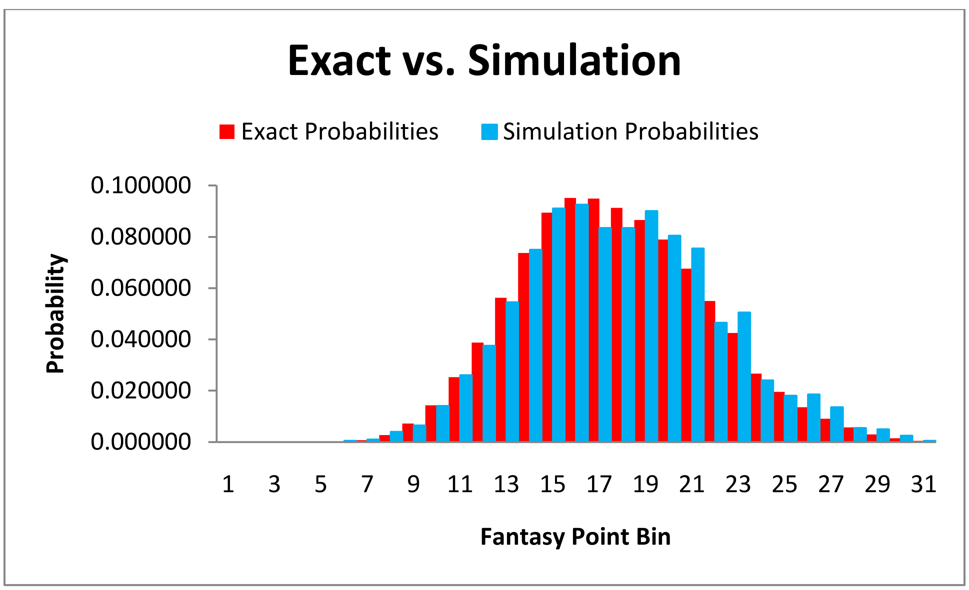


Figure 13: Exact Method vs. Simulated Method

Phase 4: Develop an Aggregate Team Fantasy Point Distribution for Each Future Week

Overview:

Calculating aggregate fantasy point player distributions is computationally intensive. Two players each with 16 bins will result in 16*16=256 convolutions of possible bins using Function 4. As the number of players grows, the aggregate team fantasy point distribution requires a very large number of computations. Thus, results are cached in the season calculator and the same combinations do not need to be recalculated each time they occur. There are four steps to build the weekly aggregate fantasy point player distributions.

I. Use Function 3 to develop the “best of” player distribution from starter and backups
II. Use Function 5 to add additional deviation to the projection distributions
III. Use Function 2 to add zero fantasy point event probability
IV. Use Function 4 to develop team distribution by aggregating starters’ distributions

I. Using Function 3 to develop the “best of” player distribution from starter and backups

The figure below show the steps used to build the “best of” distribution curve. These distribution curves represent only the standard deviation of the projections (STDpp), not the accuracy of the projections compared with actual results (STDpa).

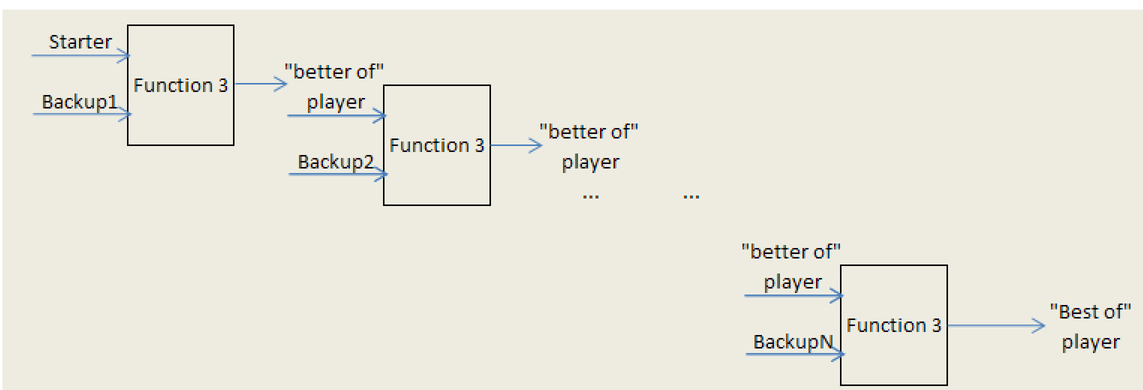


Figure 14: Developing the "Best of" Player Distribution

Function 3 is used to build the “better of” curve of two players. The resulting curve will be used again with a third player to build another “better of” curve using Function 3. This process is repeated until all starters and backups have been integrated. In the end, the curve will be a “superplayer” distribution because you have evaluated every single player (the starter and all backups) assuming they are all healthy and ready to play.

Notice that if the backup has a distribution which is centered far to the left of the starter, the backup will add little value to the “better of” distribution because the starter is the one mostly likely to be played. However, if the distributions of the backups and the starters are very close together, then the probability of scoring in the same bin will be greatly improved. Function 3 provides the mathematical method for calculating the "better of" distribution of two players for a single roster position.

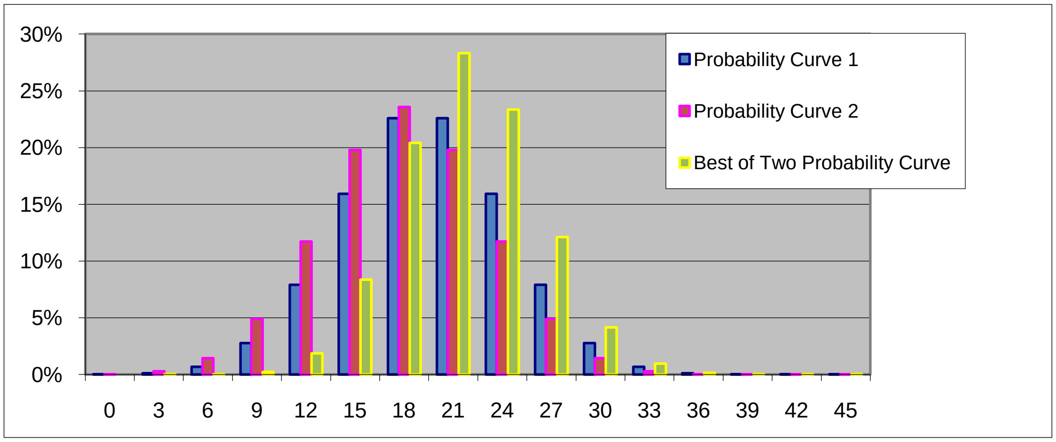


Figure 15: The "Better of" Distribution of Two Players

II. Using Function 5 to add additional deviation to the projection distributions

The figure below shows the steps used to add additional standard deviations STDpa or one-sigma to the "superplayer" distribution developed using Function 5. These steps reflect the accuracy of the projection compared with the actual result by creating normal "kernels" centered on the mid-point of each bin with STDpa as the standard deviation. The resulting distribution reflects the variations of projections compared with actual results.

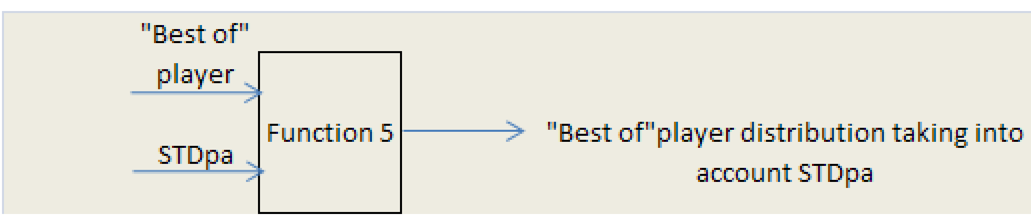


Figure 16: Adding Additional Deviation

III. Using Function 2 to add zero fantasy point event probability

The figure below shows the steps used to add the probability that injury, contract disputes, suspension, etc. will result in zero fantasy points for all players for a roster position to the "superplayer with STDpa" distribution. Use Function 2 to develop the final distribution by adjusting the probabilities of all bins and adding the probability of the zero-point event in the first bin.

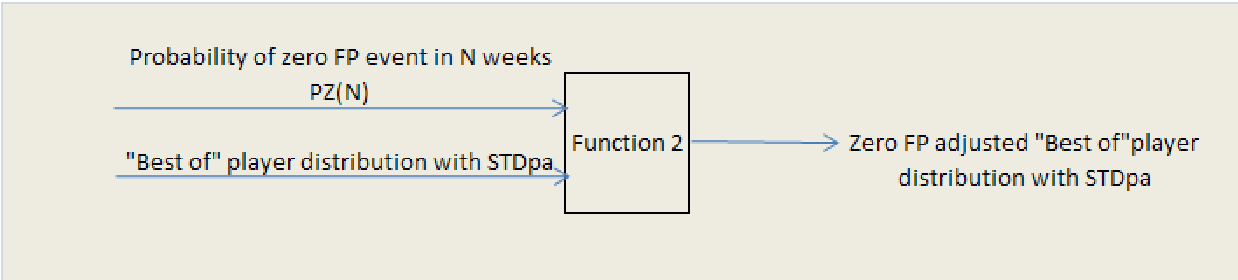


Figure 17: Adjusting Probability to Include the "Zero Point Event"

Assume the probability of achieving zero fantasy points is constant for all weeks. Denote PZ (1) as the probability of a zero fantasy point event one week from the current week. To calculate the probability of zero fantasy point event N weeks out from the current week, PZ (N), the formula is:

$$PZ(N) = 1 - (1 - PZ(1))^N$$

The probability of all players having a zero fantasy point event in week N, denoted by CPZ (N), is calculated as follows:

*CPZ(N) = PZ(N)player1 * PZ(N)player2 * PZ(N)player3 * PZ(N)player4 * ...*

Using Function 2, the final probability distribution can be calculated for the position.

The following table provides the PZ (1) values for each position:

Position	PZ(1)
Quarterback (QB)	0.012886552
Running back (RB)	0.01115234
Wide receiver (WR)	0.004208317
Tight End (TE)	0.015227113
Place Kicker (PK)	0.011095167
Defensive Team (DT)	0
Defensive Line (DL)	0.0131973
Linebacker (LB)	0.011095167
Defensive back (DB)	0.0131973

Figure 18: "Zero Point Event" Probabilities per Position

IV. Using Function 4 to develop a team distribution using aggregating starters' distributions

The figure below shows the steps used to build weekly team distributions. Using Function 4, which requires two distributions with min and max bin values, the overall team distribution can be calculated recursively.

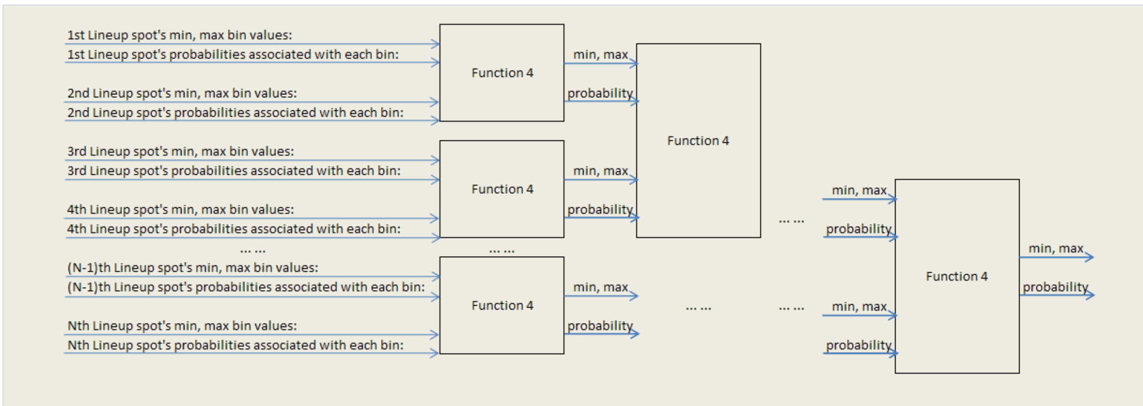


Figure 19: Developing the Team Distribution

The min and max bin values of the team aggregate distribution can be calculated as follows:

Inputs:		Min	Max	Mean	Max-Mean	× 2^(-1/2)	
Two players' original min and max bin values	Player 1	40	120	80	40	28.28427125	
	Player 2	11.7	68.3	40	28.3	20.01112191	
Output:						New Min	New Max
Aggregate of two players' min and max bin values	sum(means)=	120		sum(max distance to mean)=	48.29539316	120-48.29	120+48.29

Figure 20: Calculating the Min and Max Bin Value of the Team Distribution

Phase 5: Develop End-of-Season Win/Loss Distributions for Each Team (Using Probability Trees)

Overview:

Phase 5 combines multiple processes to develop a win/loss distribution for each team in the league. The Season Calculator begins by inputting the output curve of Function 5 (Team Aggregate with additional deviation) into Function 6. Function 6 will output the probability a certain team will win their matchup in a certain week.

Example:

For instance, Function 6 may output the following numbers for a certain matchup schedule for Team A, where p is the probability of winning that matchup, and q is the probability of losing:

Team A	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
p =	0.8	0.6	0.75	0.34	0.55	0.25	0.88	0.72	0.63	0.49	0.7	0.39	0.52
q =	0.2	0.4	0.25	0.66	0.45	0.75	0.12	0.28	0.37	0.51	0.3	0.61	0.48

Figure 21: A Simulated Season for Team A

This data implies that Team A has an 80% chance of winning their Week 1 matchup. Conversely, whoever Team A plays during Week 1 has a 20% chance of winning their matchup. Function 6 performs these calculations for each remaining matchup for the rest of the season. Using this data, a team-specific probability tree is then constructed which details the chance a certain team has of obtaining each possible win/loss record. This probability tree can be constructed at any point in the season, and is assembled for each team in the league. The probability tree for Team A at the beginning of the season would look like this:

Team A	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
p =	0.3	0.4	0.6	0.6	0.2	0.6	0.4	0.3	0.8	0.6	0.5	0.4	0.3
q =	0.7	0.6	0.4	0.4	0.8	0.4	0.6	0.7	0.2	0.4	0.5	0.6	0.7
													13--0
													1.79159E-05
												12--0	
												5.97197E-05	12--1
											11--0		0.000347867
											0.000149299	11--1	
										10--0		0.001020211	11--2
										0.000298598	10--1		0.00300166
									9--0		0.002326579	10--2	
									0.000497664	9--1		0.007625042	10--3
								8--0		0.00435456	9--2		0.015249545
								0.00062208	8--1		0.015572736	9--3	
							7--0		0.006925824	8--2		0.033040051	9--4
							0.0020736	7--1		0.026790912	8--3		0.050894957
						6--0		0.00850176	7--2		0.059241024	8--4	
						0.00518	6--1		0.040034304	7--3		0.092556403	8--5
					5--0		0.0235008	6--2		0.091691136	7--4		0.117799668
					0.00864	5--1		0.04791744	6--3		0.142529472	7--5	
				4--0		0.05098	5--2		0.126129024	6--4		0.176700618	7--6
				0.0432	4--1		0.1048896	5--3		0.193367808	6--5		0.194420449
			3--0		0.0792	4--2		0.14568192	5--4		0.227957338	6--6	
			0.072	3--1		0.18576	4--3		0.238193664	5--5		0.23576672	6--7
		2--0		0.2232	3--2		0.240864	4--4		0.262546867	5--6		0.231473367
		0.12	2--1		0.2568	3--3		0.2613216	4--5		0.247480794	5--7	
	1--0		0.324	2--2		0.32352	3--4		0.278782336	4--6		0.221455542	5--8
	0.3	1--1		0.3912	2--3		0.309056	3--5		0.23241472	4--7		0.19858351
0--0		0.46	1--2		0.368	2--4		0.28314752	3--6		0.182417664	4--8	
	0--1		0.436	1--3		0.28736	2--5		0.201502976	3--7		0.145215437	4--9
	0.7	0--2		0.2752	1--4		0.2226944	2--6		0.132420608	3--8		0.121068518
		0.42	0--3		0.2336	1--5		0.18109184	2--7		0.089412096	3--9	
			0.168	0--4		0.1257	1--6		0.086365696	2--8		0.064725709	3--10
				0.0672	0--5		0.0840192	1--7		0.046403584	2--9		0.050875972
					0.05376	0--6		0.06268416	1--8		0.027696128	2--10	
						0.0215	0--7		0.019762176	1--9		0.018559918	2--11
							0.0129024	0--8		0.008988672	1--10		0.013909303
								0.00903168	0--9		0.004855603	1--11	
									0.001806336	0--10		0.003057869	1--12
										0.000722534	0--11		0.002205536
											0.000361267	0--12	
												0.00021676	0--13
													0.000151732
Sum:	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 22: Probability "Tree" for Team A's Simulated Season

This tree details all of Team A's possible outcomes for the season. Each of these outcomes has an associated probability. Given the win/loss probabilities calculated by Function 6, each of Team A's records and their probabilities are tabulated above. If p_1 is the probability of a win in Week 1, then the formula for determining an undefeated record is simply a product of all the p_j's (inclusive from 1 to 13 in this example) where j is the number of the week. This formula is shown in the above table, and is used under the assumption that no previous games have been played. Once a team loses however, they can no longer obtain an undefeated record, so the entire "undefeated" string is unreachable (the converse applies here as well; once you have won a game, the lowest branch of the tree leading to an 0-13 record is also unreachable).

To obtain the probability of achieving any other record, one must analyze the path necessary to obtain that record. For instance, to obtain a record of 1-1, it is necessary to account for the two ways that record could happen: a win with a prior 0-1 record, and a loss with a prior 1-0 record. We multiply the probability of a 0-1 record with the probability of an upcoming win, and add that to the product of the probability of a 1-0 record and an upcoming loss.

To find the probability of any other record (any other combination of wins and losses), a similar method can be used. To find the probability of a 2-2 record for instance, we multiply the probability of a 1-2 record with the probability of an upcoming win, and add that to the product of the probability of a 2-1 record with the probability of an upcoming loss. These three methods can be used to completely define all of the probabilities for all of their corresponding records.

Once all probabilities have been defined from the data, a distribution can be developed and output:

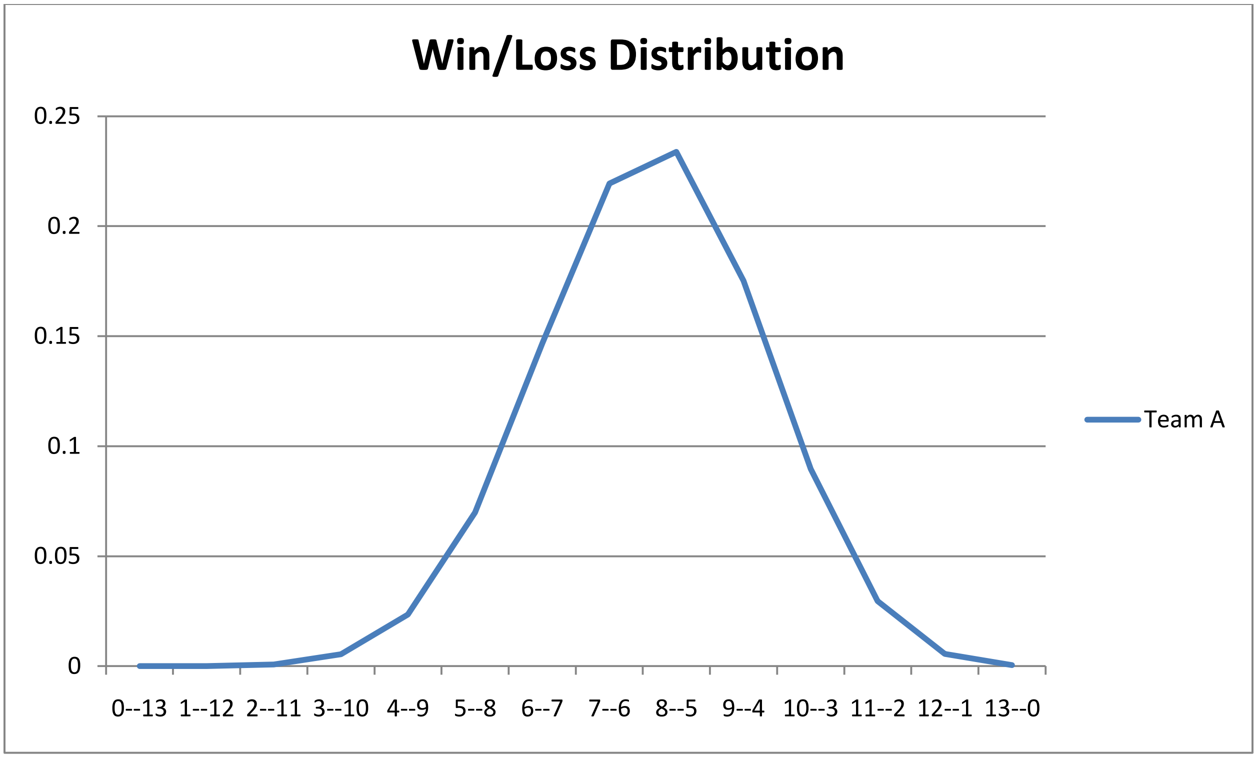


Figure 23: The Win/Loss Distribution for Team A

Similarly, a new probability tree can be developed once a team is in season. The corresponding F6 probabilities are updated given the actual matchup data, with a 1 implying a 100% chance that the outcome happened. For a team with a 6-2 record in week 8, (with wins in weeks 1, 2, 4, 5, 6, 8 and losses in weeks 3 and 7) the probability tree is updated as follows:

Team A	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
p =	1	1	0	1	1	1	0	1	0.8	0.6	0.5	0.4	0.3
q =	0	0	1	0	0	0	1	0	0.2	0.4	0.5	0.6	0.7
													13--0
													0
												12--0	
												0	12--1
											11--0		0
											0	11--1	
										10--0		0	11--2
										0	10--1		0.0288
									9--0		0	10--2	
									0	9--1		0.096	10--3
								8--0		0	9--2		0.1656
								0	8--1		0.24	9--3	
							7--0		0	8--2		0.328	9--4
							0	7--1		0.48	8--3		0.3436
						6--0		0	7--2		0.46	8--4	
						0	6--1		0.8	7--3		0.38	8--5
					5--0		0	6--2		0.44	7--4		0.3176
					0	5--1		1	6--3		0.26	7--5	
				4--0		1	5--2		0.2	6--4		0.172	7--6
				0	4--1		1	5--3		0.08	6--5		0.1276
			3--0		1	4--2		0	5--4		0.04	6--6	
			0	3--1		0	4--3		0	5--5		0.024	6--7
		2--0		1	3--2		0	4--4		0	5--6		0.0168
		1	2--1		0	3--3		0	4--5		0	5--7	
	1--0		1	2--2		0	3--4		0	4--6		0	5--8
	1	1--1		0	2--3		0	3--5		0	4--7		0
0--0		0	1--2		0	2--4		0	3--6		0	4--8	
	0--1		0	1--3		0	2--5		0	3--7		0	4--9
	0	0--2		0	1--4		0	2--6		0	3--8		0
		0	0--3		0	1--5		0	2--7		0	3--9	
			0	0--4		0	1--6		0	2--8		0	3--10
				0	0--5		0	1--7		0	2--9		0
					0	0--6		0	1--8		0	2--10	
						0	0--7		0	1--9		0	2--11
							0	0--8		0	1--10		0
								0	0--9		0	1--11	
									0	0--10		0	1--12
										0	0--11		0
											0	0--12	
												0	0--13
													0
Sum:	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 24: Probability "Tree" for Team A's Simulated Season after Week 8

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

The yellow path in the tree above shows the path that Team A took to reach a 6-2 record. All current unobtainable records which have been mathematically eliminated are shown in lilac. This leaves a new tree starting from a 6-2 record, with updated probabilities shown (given the same F6 data for the rest of the season).

Since the updated tree has a new probability space (as more than half of the possible ending records have been eliminated), the win / loss probability distribution has a new curve as well:

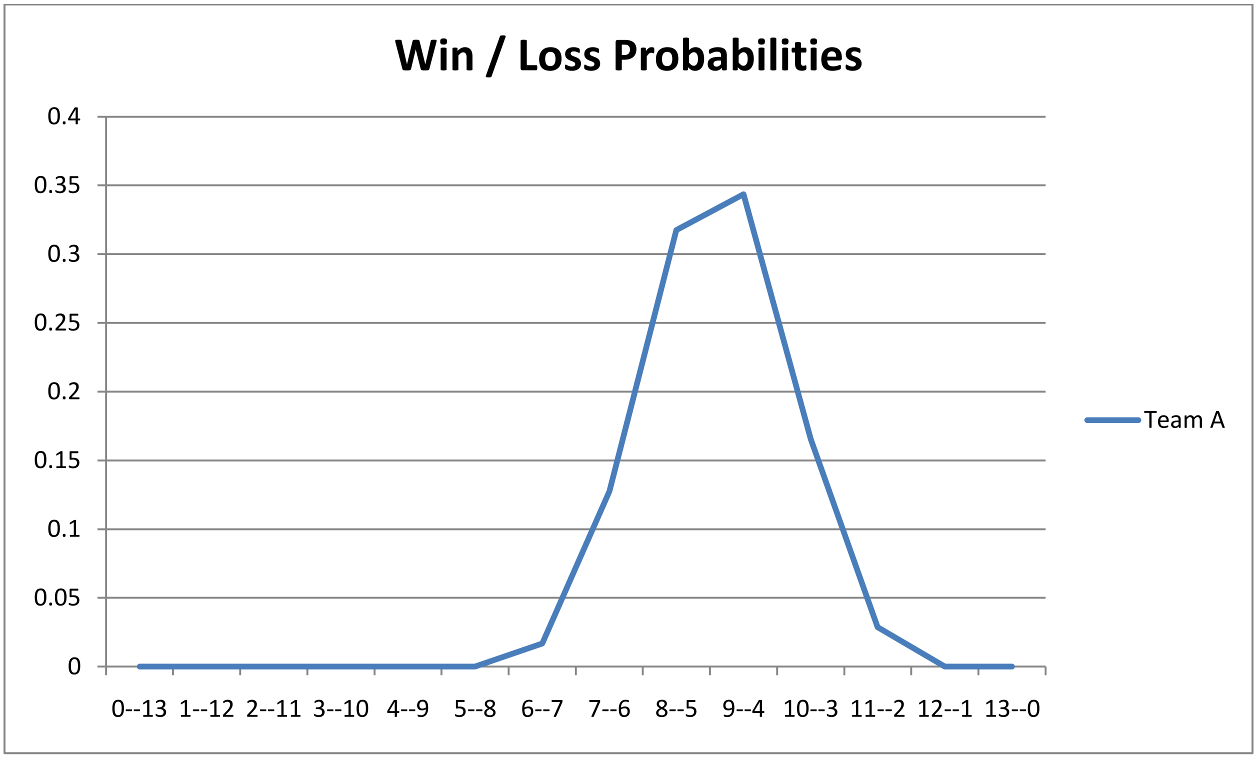


Figure 25: The 8-Week Adjusted Win/Loss Distribution for Team A

This new curve is mathematically appropriate since it is now centered on the more likely records, the unobtainable records have been eliminated, and the area under the curve still sums to 1. Note, however, that in developing these win/loss distribution curves for each team, an assumption of independence of events has been made – but in reality, if Team A goes undefeated, no other team in the league can achieve that same result, and so there is some dependence between the outcomes. This point is discussed more later in the report.

Phase 6: Determine Each Team's Percentage Chance of Making the Playoffs

Overview:

In order to properly determine a team's chances of winning their playoffs and therefore their league, it is necessary to first identify what their chance of going to the playoffs is.

Purpose:

The purpose of this phase is to determine each team's chance of making into the playoffs. Due to differences in league rules, there may or may not be divisions within the league, and the number of teams going to the playoffs will vary from league to league. In the current version of the season calculator seeding does not take into account divisions, however with a simple modification to the current method, it can.

Requirements:

I. All team's win/loss record distributions

Process:

See Function 7 for more information

Example:

See Function 7 for more information

Alternative Proposed Method:

We will demonstrate our alternative proposed method using an example. Suppose there are 8 teams: A, B, C, D, E, F, G, and H. Those eight teams are broken up into two divisions with Division 1 containing teams A-D and Division 2 containing teams E-H. In this example, three teams will be going to the playoffs, the top team from each division and a wildcard.

Assuming a schedule as follows (where a 1 signifies a win, 0 a loss) a possible combinations of wins and losses may be:

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
AvB	AvC	AvD	AvE	AvF	AvG	AvH
CvD	BvD	BvC	BvF	BvG	BvH	BvE
EvF	EvG	EvH	CvG	CvH	CvE	CvF
GvH	FvH	FvG	DvH	DvE	DvF	DvG
10101010	11001100	11001100	11110000	10101010	10011001	10011100

Figure 26: Schedule and Win/Loss Record

We can sum up total wins for each team for that particular combination:

A	B	C	D	E	F	G	H
7	3	3	3	6	3	2	1

Figure 27: Number of Wins of 8 Teams

Then we can say (where p is the probability of the specific combination of win/loss happening above given the schedule):

p	Division I				Division II			
Team	A	B	C	D	E	F	G	H
Wins	7	3	3	3	6	3	2	1
Seeding	1	3	T4	T4	2	T4	7	8

Figure 28: Seeding of 8 Teams

Since A has the most wins overall, it is seed 1. Since E has the most wins out of the remaining teams not in division 1, it is seed 2. Since B has 3 wins, and all of its wins came over C, D, and F (as seen above in Weeks 2, 3, and 4), it is seed 3. Since C, D, and F all have 3 wins, but cyclically beat one another, their seeding comes down to the probability that they score more fantasy points overall than the other two teams. G and H's seeding should now be obvious.

Since we have the probability that a team wins its matchup each week, we can determine a number for p (the specific combination happening) by multiplying the probabilities of each weekly combination happening by one another.

For instance, for a smaller case with teams A, B, C, D involving only 3 weeks, with the same matchups starting at week 1, and the following probabilities that the team will win its matchup on a given week:

Week	1	2	3
A	0.6	0.5	0.6
B	0.4	0.6	0.5
C	0.7	0.5	0.5
D	0.3	0.4	0.4

Figure 29: Probability of Winning for Each Team

It follows that:

Week 1	A	B	C	D	
	1	0	1	0	0.42
	1	0	0	1	0.18
	0	1	1	0	0.28
	0	1	0	1	0.12
					1

Figure 30: Probability of Winning in Week 1

Where .42 = .6*.7 represents the scenario in which A and C win their matchups, etc.

Week 2	A	B	C	D	
	1	0	0	1	0.2
	1	1	0	0	0.3
	0	1	1	0	0.3
	0	0	1	1	0.2
					1

Figure 31: Probability of Winning in Week 2

Week 3	A	B	C	D	
	1	1	0	0	0.3
	1	0	1	0	0.3
	0	1	0	1	0.2
	0	0	1	1	0.2
					1

Figure 32: Probability of Winning in Week 3

So, then the probability that the particular combination (1010, 1001, 1100) happens (where the commas separate weeks) is .0252 = .42*.2*.3

Note that there would be $(2*2)^3 = 64$ combinations for this example.

So back to our original case of:

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
AvB	AvC	AvD	AvE	AvF	AvG	AvH
CvD	BvD	BvC	BvF	BvG	BvH	BvE
EvF	EvG	EvH	CvG	CvH	CvE	CvF
GvH	FvH	FvG	DvH	DvE	DvF	DvG
10101010	11001100	11001100	11110000	10101010	10011001	10011100

Figure 33: Winning String of All Teams

With:

p	Division I				Division II			
Team	A	B	C	D	E	F	G	H
Wins	7	3	3	3	6	3	2	1
Seeding	1	3	T4	T4	2	T4	7	8

Figure 34: Accounting for Divisions

And if p_{A1}=probability that Team A wins its matchup week 1, etc., then the probability this combo happens (p) is:

$$p=(p_{A1}*p_{C1}*p_{E1}*p_{G1})*(p_{A2}*p_{B2}*p_{E2}*p_{F2})*(p_{A3}*p_{B3}*p_{E3}*p_{F3})*(p_{A4}*p_{B4}*p_{C4}*p_{D4})*(p_{A5}*p_{C5}*p_{E5}*p_{G5})*(p_{A6}*p_{D6}*p_{E6}*p_{H6})*(p_{A7}*p_{D7}*p_{E7}*p_{F7})$$

If we arbitrarily assume that C has a 60% to score the most fantasy points overall comparative to D and F, D has a 30% chance to score the most fantasy points overall comparative to C and F, and F has a 10% chance to score the most fantasy points overall comparative to C and D (taking care of 4th place), C has a 25% chance to score the second most fantasy points overall comparative to D and F, D has a 25% chance to score the second most fantasy points overall comparative to C and F, and F has a 50% chance to score the second most fantasy points overall comparative to C and D, then that leaves us with a 15% chance that C scores the fewest overall fantasy points comparative to D and F, a 45% chance that D scores the fewest overall fantasy points comparative to C and F, and a 40% chance that F scores the fewest overall fantasy points comparative to C and D. We can then build a table that looks like this:

Seeding	A	B	C	D	E	F	G	H
1	p							
2					p			
3		p						
4			.6*p	.3*p		.1*p		
5			.25*p	.25*p		.5*p		
6			.15*p	.45*p		.4*p		
7							p	
8								p

Figure 35: Probability Table

So for this example, after considering every combination (for this case it would be $(4*2)^7$=2,097,152 combinations) and adding in each p for each combination to the table, we would have the marginal probabilities that A is seed 1 through seed 8, B is seed 1 through seed 8, etc. Specifically, this section (below) tells us two things:

Seeding	A	B	C	D	E	F	G	H
1	.09							
2	.04							
3	.02							

Figure 36: Seeding Table

The probabilities associated with each team being a specific seed (1 through 3) is the percentage chance they have of obtaining that seed comparative to other teams. For instance, if A has .09 probability of being seed 1 (after summing all p's where A ranks 1st), then A's chance of being the 1st seed comparative to the other teams in the league is 9%.

We can also say that the probability that a team makes the playoffs is the sum of the probabilities that they obtain either seed 1, 2, or 3 (if 3 teams go to the playoffs). For instance, the probability that A makes it to the playoffs comparative to the other teams is .15=.09+.04+.02. Therefore A has a 15% chance of making it to the playoffs.

We believe the above alternative proposed method will be completely accurate, but may be impractical due to its intense computational requirements; therefore, the current method is detailed in Function 7.

Phase 7: Develop Playoff Win/Loss Distribution for Each Team (Using Probability Trees)

Overview:

Phase 7 focuses on the matchups a team may face during the playoffs, and their chances of winning these matchups. Phase 7 uses the data output from Phase 6: the probabilities a team has of obtaining certain seeds. The Season Calculator builds multiple playoff possibility trees based off of the seeding data obtained in Phase 6. A win/loss distribution of playoff records can then be created by summing the probabilities of obtaining certain records.

Purpose:

The purpose of this phase is to develop playoff win/loss distribution given each team's seeding probabilities from Phase 6.

Requirements:

I. Each team's seeding probabilities from Phase 6 and Function 7
II. Each team's fantasy point distribution for the playoffs

Example:

For instance, assume that four teams go into the playoffs in a twelve team league (with no divisions). Team 1 then has a probability of being Seed 1, Seed 2, Seed 3, Seed 4 or none. The probability that Team 1 wins the first round of the playoffs is broken down into many parts. First, Team 1 can be Seed 1. To get the probability of winning the first game in that bracket (Seed 1 vs. Seed 4), two summations must be added (cases in which Team 1 is either Seed 1 or Seed 4 as their opponent would produce the same matchup within the bracket regardless of being first or last seed). The first summation would be as follows (where the probability that Team 1 beats Team i is given by Function 6):

$$\sum_{i=2}^{12} Pr_{Ti\ is\ s4} * Pr_{T1\ beats\ Ti} * Pr_{T1\ is\ S1} * \frac{\sum_{j=1}^{12} Pr_{Tj\ is\ S1}}{\sum_{j=2}^{12} Pr_{Tj\ is\ S1}}$$

The second summation would be the same, except that Team 1 is now Seed 4 and they must play Team i as Seed 1 (the quotient would change to Seed 4 as well). These two summations of probabilities must be added together to produce the probability that Team 1 wins the game if it is in that bracket. Then, the same process can be followed to determine Team 1's chance of winning the first game if it is in the Seed 2 vs. Seed 3 bracket. Adding these two probabilities together sums up the total amount of ways that Team 1 can be in the first round of the playoffs and win the matchup. If there were more first round brackets, they are added in at this step.

The process continues in similar fashion, except instead of using the probability that Team 1 is Seed N, the probability that they win their first matchup is used instead. The probability that Team i wins their first matchup is used likewise. It should be noted that both Team 1 and Team i are unable to face either their or their opponent's first round adversary. This process follows throughout all rounds of the playoffs up through the final. The probability of winning the championship requires just one more iteration of the above formula and would include all possible scenarios of both how to get to the championship and win it.

This method simply extends each combination further using probability trees as defined in Phase 5. For instance, given the combination:

p	Division I				Division II			
Team	A	B	C	D	E	F	G	H
Wins	7	3	3	3	6	3	2	1
Seeding	1	3	5	4	2	6	7	8

Figure 37: Seeding of 8 Teams

Multiple combinations for the playoffs can be assembled. The playoffs start with A vs. D, and E vs. B. For the first combination, we build a tree where A plays against D, then E (random win / loss probabilities have been assigned):

	vs. D	vs. E
Team A	Round 1	Round 2
p =	0.3	0.6
q =	0.7	0.4

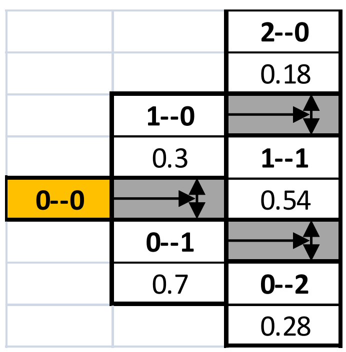


Figure 38: First Week Matchup

Therefore, the probability that A wins the playoffs given this combination is 18%. The other combination must be considered as well (where A plays B in the finals). This combination would be as follows (with random probability assignment once again):

	vs. D	vs. B
Team A	Round 1	Round 2
p =	0.3	0.8
q =	0.7	0.2

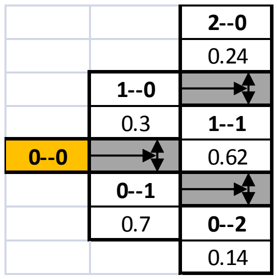


Figure 39: Probability of Winning in Second Week

Phase 8: Output Percentage Chance of Winning the Championship

Overview:

Phase 8 uses the data created in the playoff probability tree assembled in Phase 7 to determine a team's overall chance of winning the championship. This data includes all possible combinations of win/loss records which result in the team reaching the playoffs, progressing within the playoffs to the championship game, and winning the championship. As shown in Phase 7, the probability of this occurrence is a product of the likelihood of the team reaching the finals and the probability of the team winning the last match. The probability that the team reaches the finals is retrospective in the sense that it takes into account the probabilities of both winning and being in previous playoff matchups. A summation of these probabilities (one for each combination of teams that can be faced) determines a team's chances of winning the championship. This final cumulative percentage is the output of Phase 8 and the final result of the Season Calculator.

Functions Used in the Season Calculator

The previously described eight phases make use of the following 10 functions, and other mathematical processes:

Function 1: Histograms of a Normal Distribution

Overview:

Function 1 transforms a normal distribution into a discrete probability histogram.

Purpose:

The purpose of this function is to convert a normal distribution with a mean and a standard deviation to a histogram. The output of Function 1 is probabilities, Data Type 1

Requirements:

I. A mean
II. A standard deviation
III. A max bin value

IV. 16 bins as separators, Data Type 2

Example:

Suppose we are in Week 2, Tom Brady has a mean of 22.5 fantasy points in Week 5, with a one-sigma value of 6. The max bin value is 45 fantasy points for the quarterback position; therefore the probability bins (DT2) range from 0 to 45.

Question:

How do we create a histogram of Tom Brady's distributions using a normal distribution with a mean of 22.5 and a standard deviation of 6?

Mathematical Background:

We are given a continuous random variable X that follows a normal distribution with mean μ and a standard deviation of σ. We restrict X to range from 0 to the given max bin value with 16 bins. We can calculate the incremental bin values by:

$$\text{incremental bin value} = \text{max bin value} \div (16 - 1)$$

$$\text{incremental bin value} = 45 \div 15$$

$$\text{incremental bin value} = 3$$

Therefore, X can only take on values in 16 bins:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(-∞,3]	(3,6]	(6,9]	(9,12]	(12,15]	(15,18]	(18,21]	(21,24]	(24,27]	(27,30]	(30,33]	(33,36]	(36,39]	(39,42]	(42,45]	(45,∞)

Figure 40: Fantasy Point Bins

Instead of "chopping off" a normal curve and distributing probabilities at 3, 6, 9, 12 and up to 45, we "chop off" probabilities at the midpoint of those bins. Below is a step-by-step guide on how to do this:

Step 1: Generage a Culmulative Normal Distribution with a mean and a standard deviation																
μ=	22.5	σ=	6													
Bins	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	(-∞,3]	(3,6]	(6,9]	(9,12]	(12,15]	(15,18]	(18,21]	(21,24]	(24,27]	(27,30]	(30,33]	(33,36]	(36,39]	(39,42]	(42,45]	(45,∞)
Midpoint	1.5	4.5	7.5	10.5	13.5	16.5	19.5	22.5	25.5	28.5	31.5	34.5	37.5	40.5	43.5	46.5
F(X≤Midpoint	0%	0%	1%	2%	7%	16%	31%	50%	69%	84%	93%	98%	99%	100%	100%	100%
Step 2: Calculate Probabilities belong to each bin by taking the difference of F(X≤Midpoint) except the first one																
Bins	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	(-∞,3]	(3,6]	(6,9]	(9,12]	(12,15]	(15,18]	(18,21]	(21,24]	(24,27]	(27,30]	(30,33]	(33,36]	(36,39]	(39,42]	(42,45]	(45,∞)
Probabilities	0%	0%	0%	2%	4%	9%	15%	19%	19%	15%	9%	4%	2%	0%	0%	0%

Figure 41: Distribution of Fantasy Points

Here is what the histogram will look like:

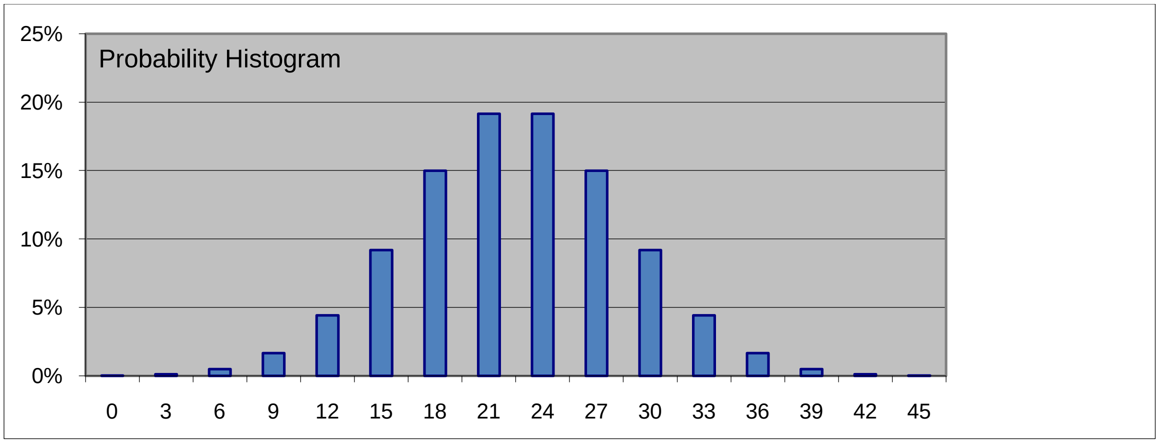


Figure 42: Graph of the Distribution

How it can be used in Function 1:

Back to the example, given a normal distribution with a mean of 22.5 and a standard deviation of 6, using Function 1, we can convert it into a probability histogram. Here is a summary:

Point	Mid Point	F(X≤Mid Point)	**Bin Value**	**Probability**	mean	22.5
0	1.5	0%	**(-∞,3]**	**0%**	σ	6
3	4.5	0%	**(3,6]**	**0%**		
6	7.5	1%	**(6,9]**	**0%**		
9	10.5	2%	**(9,12]**	**2%**		
12	13.5	7%	**(12,15]**	**4%**		
15	16.5	16%	**(15,18]**	**9%**		
18	19.5	31%	**(18,21]**	**15%**		
21	22.5	50%	**(21,24]**	**19%**		
24	25.5	69%	**(24,27]**	**19%**		
27	28.5	84%	**(27,30]**	**15%**		
30	31.5	93%	**(30,33]**	**9%**		
33	34.5	98%	**(33,36]**	**4%**		
36	37.5	99%	**(36,39]**	**2%**		
39	40.5	100%	**(39,42]**	**0%**		
42	43.5	100%	**(42,45]**	**0%**		
45	46.5	100%	**(45,∞)**	**0%**		

Figure 43: Distribution of Fantasy Points

Function 2: Adding the Probability of a Zero-Fantasy-Point Event

Overview:

Adjust the probability of the first bin and change the rest of the bins accordingly

Purpose:

The purpose of this function is to change the probabilities of other bins proportionally and add probability to the first bin so that the sum of all probabilities is still equal to 1. This function is used to add the probability of a zero-fantasy-point event. The output of Function 2 will still be a histogram.

Requirements:

I. The probability of a zero-fantasy-point event
II. A probability distribution represented as Data Type 1

Example:

Suppose the team aggregate distribution after using Function 5 is as follows: the team has a mean of 22.5 fantasy points in Week 5, with a one-sigma value of 6.

Bin Value	Probability	mean	22.5
(-∞,3]	0%	σ	6
(3,6]	0%		
(6,9]	0%		
(9,12]	2%		
(12,15]	4%		
(15,18]	9%		
(18,21]	15%		
(21,24]	19%		
(24,27]	19%		
(27,30]	15%		
(30,33]	9%		
(33,36]	4%		
(36,39]	2%		
(39,42]	0%		
(42,45]	0%		
(45,∞)	0%		

Figure 44: Distribution of Fantasy Points

The projections do not yet reflect the possibility that any players on the team could be injured (as mentioned in Phase 4 part C) and therefore will not be playing in Week 5. This will add more probability to the zero-fantasy-point event.

Question:

Suppose the probability of being in the first bin becomes 5%. How do we make adjustments so that the probability of scoring zero fantasy points increases and the overall probabilities are not affected?

Mathematical Background:

We are given a discrete distribution of random variable X and the probability of X being a specific value or range has been changed. We can break the random variable X into the sum of random variables of Y and Z, where Y represents a point-mass distribution where X's value or range has changed, and Z represents the remaining values that X can take on. Let X's probability of being that value or range be p. Then the new distribution of X will be:

$$X = \begin{cases} Y * p \\ Z * (1 - p) \end{cases}$$

How it can be used in Function 2:

Back to the example, multiply every other probability by (1-5%), or 95% and set the first probability to 5%. Here is a summary:

ColumnA	ColumnB	multiply ColumnB by 95%		
Bin Value	**Probability**	New Probability	mean	22.5
(-∞,3]	**0%**	5%	σ	6
(3,6]	**0%**	0%		
(6,9]	**0%**	0%		
(9,12]	**2%**	2%		
(12,15]	**4%**	4%		
(15,18]	**9%**	9%		
(18,21]	**15%**	14%		
(21,24]	**19%**	18%		
(24,27]	**19%**	18%		
(27,30]	**15%**	14%		
(30,33]	**9%**	9%		
(33,36]	**4%**	4%		
(36,39]	**2%**	2%		
(39,42]	**0%**	0%		
(42,45]	**0%**	0%		
(45,∞)	**0%**	0%		

Figure 45: Resulting Distribution

Function 3: "Better of" Distribution of X and Y

Overview:

Find Z=max(X, Y)

Purpose:

The purpose of this function is to develop a distribution for the "better of" two players in future weeks, given their individual projections. At the beginning of each week during the season, the team manager can freely choose any one of the two players from his/her roster and put him in the starting line-up. The "better" player will be the player who, at the beginning of the week in question, has the better projection for that week. Until the week arrives, the "better" player is unknown.

Requirements:

I. Two independent (normal) distributions

Example:

Suppose you have two quarterbacks, Tom Brady and Peyton Manning on your roster, but only one will play in any given week. Let's say we are in Week 2 and we are interested in the Week 12 projected fantasy points from the quarterback position. The mean and one-sigma values of each player's Week 12 projections are as follows:

Week 12 from Week 2	Tom Brady	Peyton Manning
mean	19.5	18
1-sigma	5	5

Figure 46: TB and PM's Mean and STD

As of Week 2, Tom Brady is projected to have a mean of 19.5 fantasy points and a standard deviation of 5 in Week 12; Peyton Manning is projected to have a mean of 18 fantasy points and the same standard deviation. Note, however, that by the time Week 12 arrives, those projections may have changed (due to injuries, line-up changes, etc.) and at that point Manning may have the higher projection. Function 3 develops a probability distribution that reflects this possibility.

Using Function 1, the Week 12 projected distribution for each player can be developed as follows:

	Tom Brady	Peyton Manning
mean	19.5	18
σ	5	5
Bin Value	**Probability**	**Probability**
(-∞,3]	**0%**	**0%**
(3,6]	**0%**	**0%**
(6,9]	**1%**	**1%**
(9,12]	**3%**	**5%**
(12,15]	**8%**	**12%**
(15,18]	**16%**	**20%**
(18,21]	**23%**	**24%**
(21,24]	**23%**	**20%**
(24,27]	**16%**	**12%**
(27,30]	**8%**	**5%**
(30,33]	**3%**	**1%**
(33,36]	**1%**	**0%**
(36,39]	**0%**	**0%**
(39,42]	**0%**	**0%**
(42,45]	**0%**	**0%**
(45,∞)	**0%**	**0%**

Figure 47: TB and PM's Distributions

Note: Probabilities are round off to nearest whole percentages.

Question:

How do we find the "better of" distribution Z using Tom Brady's distribution X and Peyton Manning's distribution Y? In mathematical terms, we are interested in the distribution of Z, where Z is defined as:

$$Z = max(X, Y)$$

Mathematical Background:

Given two independent random variables X and Y, each with its own probability distribution, define Z to be the maximum of X and Y.

The cumulative distribution of Z is:

$$\mathrm{Fz}\ (\mathrm{z})$$

$$= \mathrm{P}\ (\mathrm{Z} \leq \mathrm{z})$$

$$= \mathrm{P}\ \max(\mathrm{X}, \mathrm{Y}) \leq \mathrm{z})$$

$$= \mathrm{P}\ (\text{both X and Y are less or equal to z})$$

$$= \mathrm{P}\ (\mathrm{X} \leq \mathrm{z}) * \mathrm{P}(\mathrm{Y} \leq \mathrm{z}) \qquad (\text{Independence of X and Y})$$

$$= \mathrm{Fx}\ (\mathrm{z}) * \mathrm{Fy}\ (\mathrm{z}) \qquad (\text{definition of a CDF})$$

Therefore,

$$\mathrm{Fz}\ (\mathrm{z}) = \mathrm{Fx}\ (\mathrm{z}) * \mathrm{Fy}\ (\mathrm{z})$$

How it can be used in Function 3:

$$\mathrm{Fz}\ (\mathrm{z}) = \mathrm{Fx}\ (\mathrm{z}) * \mathrm{Fy}\ (\mathrm{z})$$

Let Z be the “better of” random variable, X and Y be the Week 12 projected fantasy point distributions for Tom Brady and Peyton Manning respectively. Below is a step-by-step guide that shows how to find Z given X and Y:

Step 1: Get Probability Mass Function of Random Variables X and Y:			
Bin Value	Tom Brady's Prob	Peyton Manning's Prob	
(-∞,3]	0%	0%	
(3,6]	0%	0%	
(6,9]	1%	1%	
(9,12]	3%	5%	
(12,15]	8%	12%	
(15,18]	16%	20%	
(18,21]	23%	24%	
(21,24]	23%	20%	
(24,27]	16%	12%	
(27,30]	8%	5%	
(30,33]	3%	1%	
(33,36]	1%	0%	
(36,39]	0%	0%	
(39,42]	0%	0%	
(42,45]	0%	0%	
(45,∞)	0%	0%	

Step 2: Compute Cumulative Distribution Function of Random Variables X and Y:				
Bin Value	Tom Brady's Prob	Peyton Manning's Prob		
(-∞,3]	0%	0%		
(3,6]	0%	0%		
(6,9]	1%	2%		
(9,12]	4%	7%		
(12,15]	12%	18%		
(15,18]	27%	38%		
(18,21]	50%	62%		
(21,24]	73%	82%		
(24,27]	88%	93%		
(27,30]	96%	98%		
(30,33]	99%	100%		
(33,36]	100%	100%		
(36,39]	100%	100%		
(39,42]	100%	100%		
(42,45]	100%	100%		
(45,∞)	100%	100%		

Step 3: Compute Cumulative Distribution Function of Random Variable Z:					
Random Variable Z and its distribution:					
Bin Value	Fz(z)=Prob(Z<=z)				
(-∞,3]	0%				
(3,6]	0%				
(6,9]	0%				
(9,12]	0%				
(12,15]	2%				
(15,18]	10%				
(18,21]	31%				
(21,24]	59%				
(24,27]	83%				
(27,30]	95%				
(30,33]	99%				
(33,36]	100%				
(36,39]	100%				
(39,42]	100%				
(42,45]	100%				
(45,∞)	100%				

Step 4: Compute Probability Mass Function of Random Variable Z:		
Random Variable Z and its distribution:		
Bin Value	Fz(z)=Prob(Z<=z)	Probability
(-∞,3]	0%	0%
(3,6]	0%	0%
(6,9]	0%	0%
(9,12]	0%	0%
(12,15]	2%	2%
(15,18]	10%	8%
(18,21]	31%	20%
(21,24]	59%	28%
(24,27]	83%	23%
(27,30]	95%	12%
(30,33]	99%	4%
(33,36]	100%	1%
(36,39]	100%	0%
(39,42]	100%	0%
(42,45]	100%	0%
(45,∞)	100%	0%

Figure 48: Steps 1-4 to Get "Better of" Curve

Bin Value	Fz(z)=Prob(Z<=z)	Probability
(-∞,3]	0%	0%
(3,6]	0%	0%
(6,9]	0%	0%
(9,12]	0%	0%
(12,15]	2%	2%
(15,18]	10%	8%
(18,21]	31%	20%
(21,24]	59%	28%
(24,27]	83%	23%
(27,30]	95%	12%
(30,33]	99%	4%
(33,36]	100%	1%
(36,39]	100%	0%
(39,42]	100%	0%
(42,45]	100%	0%
(45,∞)	100%	0%

Figure 49: Distribution of "Better of" Curve

Note: all probabilities are rounded to nearest whole percentages.

A graphical representation of the data is shown below. Using the data from above, two distributions X and Y are shown (in green and blue), then a third distribution Z (in red) represents the "better of" X and Y.

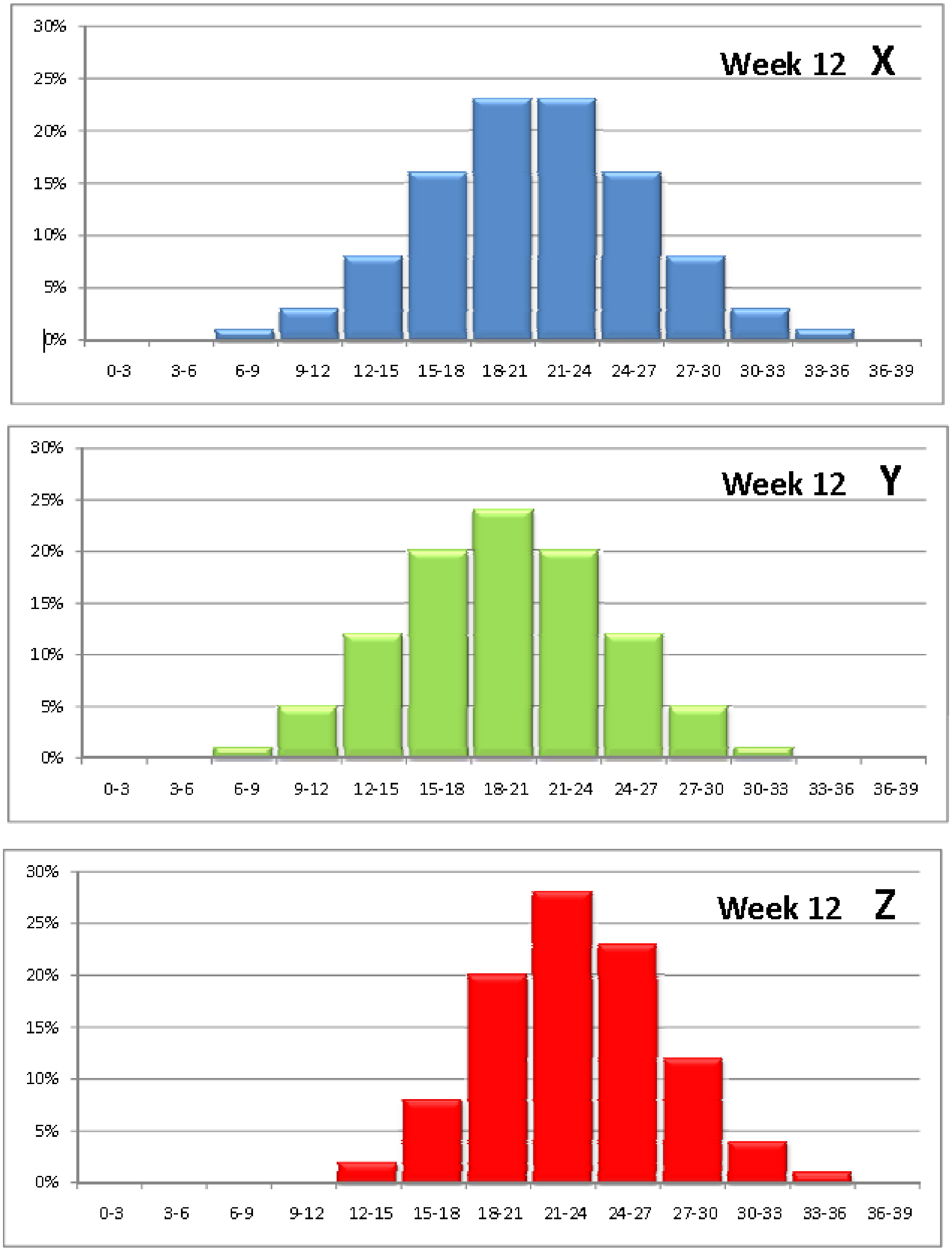


Figure 50: A Graphical Representation of Function 3

Function 4: Sum of the Two Independent Distributions

Overview:

Find Z=X+Y

Purpose:

The purpose of this function is to create a distribution that represents the sum of the two independent distributions. This function is used to develop a team aggregate fantasy point distribution. The output of Function 4 is probabilities, Data Type 1

Requirements:

I. Two independent distributions

Example:

Suppose Tom Brady's and Randy Moss's distributions are as follows:

	Tom Brady	Randy Moss
mean	19.5	25
σ	5	8
Bin Value	Probability	Probability
(-∞,3]	0%	0%
(3,6]	0%	0%
(6,9]	1%	1%
(9,12]	3%	2%
(12,15]	8%	4%
(15,18]	16%	7%
(18,21]	23%	10%
(21,24]	23%	13%
(24,27]	16%	15%
(27,30]	8%	14%
(30,33]	3%	12%
(33,36]	1%	9%
(36,39]	0%	6%
(39,42]	0%	3%
(42,45]	0%	2%
(45,∞)	0%	1%

Figure 51: Two Player Probability Distributions

Question:

How do we get a distribution of their sum?

Mathematical Background:

We are given two discrete distributions of X and Y, and each takes on 16 values. To calculate the distribution of Z:

$$Z = X + Y$$

There are 16 values of X, and for each value of X, there are 16 different values of Y, which results in 256 possible combinations of X and Y.

The calculations are a bit intense, so instead we will illustrate the process through another example: the toss of two independent, fair, 6-headed dice. What will their sum be?

Functio 4											
X=	1	2	3	4	5	6					
Prob(X=x)	17%	17%	17%	17%	17%	17%	100%				
Y=	1	2	3	4	5	6					
Prob(Y=y)	17%	17%	17%	17%	17%	17%	100%				
Since they are independent then:											
X→, Y↓	1	2	3	4	5	6					
1	3%	3%	3%	3%	3%	3%	17%				
2	3%	3%	3%	3%	3%	3%	17%				
3	3%	3%	3%	3%	3%	3%	17%				
4	3%	3%	3%	3%	3%	3%	17%				
5	3%	3%	3%	3%	3%	3%	17%				
6	3%	3%	3%	3%	3%	3%	17%				
	17%	17%	17%	17%	17%	17%	100%				
Values											
Z	1	2	3	4	5	6					
1	2	3	4	5	6	7					
2	3	4	5	6	7	8					
3	4	5	6	7	8	9					
4	5	6	7	8	9	10					
5	6	7	8	9	10	11					
6	7	8	9	10	11	12					
Probability											
Z	2	3	4	5	6	7	8	9	10	11	12
	3%	6%	8%	11%	14%	17%	14%	11%	8%	6%	3%

Figure 52: A Dice Example of the Sum of X and Y

How it can be used in Function 4:

The approach will be very similar, where you would create a 16×16 matrix and figure out the Z values and probabilities associated with each value and sum them up.

Function 5: Normal Kernel Distribution on a Histogram

Overview:

Kernel smoothing on bins where the area is equal to the probability of the bin

Purpose:

The purpose of this function is to add 16 normal "kernels" centered on the midpoint of each bin with STDpa as the standard deviation. The resulting distribution is smoother. The output of Function 5 is probabilities, Data Type 1

Requirements:

I. STDpa

II. Team aggregate distribution

Example:

Function 5: Adding deviation to a probability curve															
1-sigma:	2														
Mean:	22														
0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
0%	0%	0%	0%	0%	0%	10%	49%	36%	4%	0%	0%	0%	0%	0%	0%

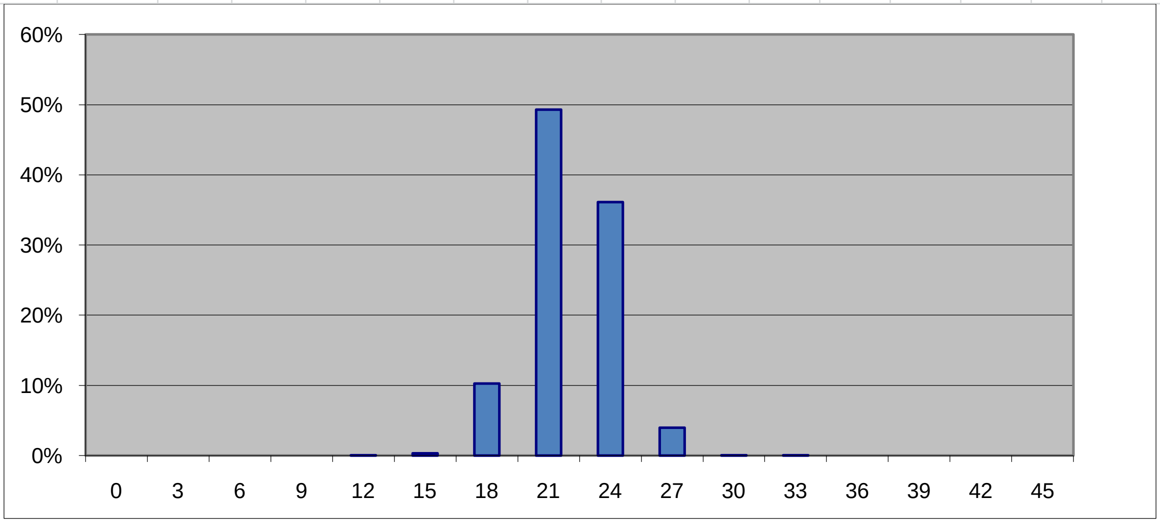


Figure 53: The Team Distribution before Additional Deviation

Create Normal Kernels on each bin with STDpa of :							7									
0	58%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
3	42%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	0%	0%	0%	0%	100%
6	26%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	0%	0%	0%	100%
9	14%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	0%	0%	100%
12	7%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	0%	100%
15	3%	4%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	0%	100%
18	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	0%	100%
21	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	0%	100%
24	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	0%	100%
27	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	4%	2%	1%	100%
30	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	4%	3%	100%
33	0%	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	8%	7%	100%
36	0%	0%	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	12%	14%	100%
39	0%	0%	0%	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	16%	26%	100%
42	0%	0%	0%	0%	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	17%	42%	100%
45	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	2%	4%	8%	12%	16%	58%	100%

Figure 54: Using Normal-Kernels to Smooth the Curve

	Multiply the previous probabilities by the corresponding bin probability															
0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
18	0%	0%	0%	1%	1%	2%	2%	2%	1%	1%	0%	0%	0%	0%	0%	0%
21	0%	0%	1%	2%	4%	6%	8%	8%	8%	6%	4%	2%	1%	0%	0%	0%
24	0%	0%	0%	1%	1%	3%	4%	6%	6%	6%	4%	3%	1%	1%	0%	0%
27	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	1%	0%	0%	0%	0%	0%
30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
33	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
36	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
39	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
42	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
45	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Figure 55: Accounting for the Probability of Being in a Certain Bin

A smoother curve results:

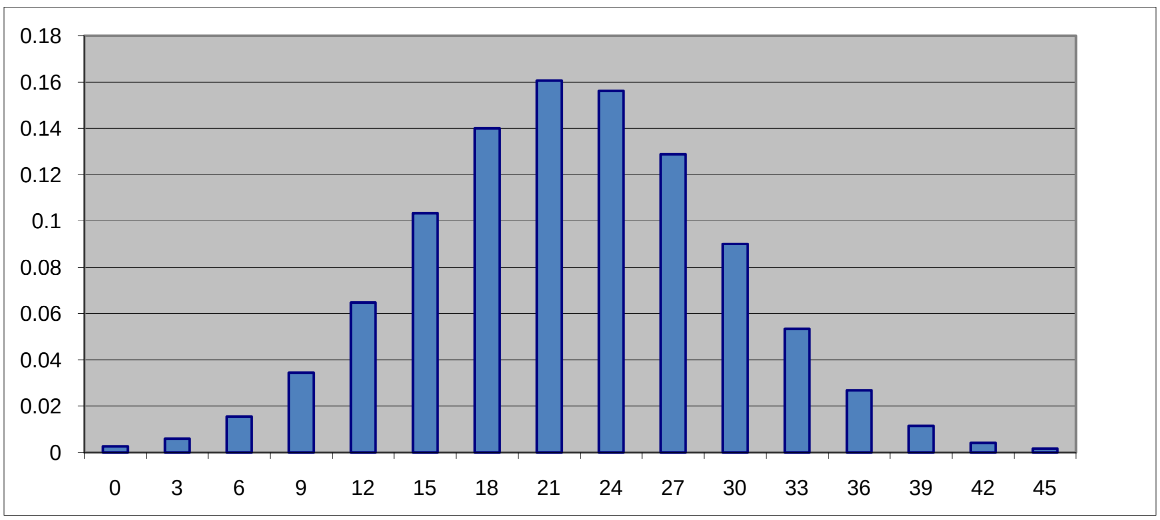


Figure 56: The Output of Function 5

Function 6: Distribution X is Greater Than Distribution Y

Overview:

Find Pr(X>Y)

Purpose:

The purpose of this function is to calculate the probability that a random variable X is greater than a random variable Y. This function is used to determine the probability that one team will achieve a higher fantasy point total than another team's. The output of Function 6 is a single probability.

Requirements:

I. Two independent distributions

Example:

Suppose Team A and Team B's distributions are as follows:

	Team A	Team B
mean	30	25
σ	5	10
Bin Value	Probability	Probability
(-∞,3]	0%	1%
(3,6]	0%	1%
(6,9]	0%	2%
(9,12]	0%	3%
(12,15]	0%	5%
(15,18]	0%	7%
(18,21]	1%	9%
(21,24]	5%	11%
(24,27]	12%	12%
(27,30]	20%	12%
(30,33]	24%	11%
(33,36]	20%	9%
(36,39]	12%	7%
(39,42]	5%	5%
(42,45]	1%	3%
(45,∞)	0%	3%

Figure 57: Distribution of A and B

Question:

How do we get the probability that Team A's total exceeds Team B's total?

Mathematical Background:

We are given two discrete distributions of X and Y, each takes on 16 values. To calculate the probability that X is greater than Y:

$$\mathrm{P}(\mathrm{X} > Y) = \sum_{\mathrm{y}} \mathrm{P}(\mathrm{X} > y \textit{ given } Y = y) \times \mathrm{P}(\mathrm{Y} = \mathrm{y})$$

Or

$$\mathrm{P}(\mathrm{X} > Y) = \sum_{\mathrm{y}} \sum_{\mathrm{x}>y} \mathrm{P}(\mathrm{X} = \mathrm{x}, \mathrm{Y} = \mathrm{y})$$

We can use a similar example to illustrate how the second formula works: the tossing of two independent, fair, 6-headed dice. What is the probability that first die achieves a higher number than the second die?

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

X=	1	2	3	4	5	6	
Prob(X=x)	17%	17%	17%	17%	17%	17%	100%
Y=	1	2	3	4	5	6	
Prob(Y=y)	17%	17%	17%	17%	17%	17%	100%
Since they are independent then: P(X,Y)=P(X=x)×P(Y=y)							
X/Y	1	2	3	4	5	6	
1	3%	3%	3%	3%	3%	3%	17%
2	3%	3%	3%	3%	3%	3%	17%
3	3%	3%	3%	3%	3%	3%	17%
4	3%	3%	3%	3%	3%	3%	17%
5	3%	3%	3%	3%	3%	3%	17%
6	3%	3%	3%	3%	3%	3%	17%
	17%	17%	17%	17%	17%	17%	100%
Since they are independent then:							
X/Y	1	2	3	4	5	6	
1	3%	3%	3%	3%	3%	3%	17%
2	3%	3%	3%	3%	3%	3%	17%
3	3%	3%	3%	3%	3%	3%	17%
4	3%	3%	3%	3%	3%	3%	17%
5	3%	3%	3%	3%	3%	3%	17%
6	3%	3%	3%	3%	3%	3%	17%
	17%	17%	17%	17%	17%	17%	100%
P(X>Y)=sum of the yellow probabilities:							
42%							

Figure 58: An Example of the Probability that One Die Rolls Higher than the Other

How it could be used in Function 6:

The approach will be very similar to the dice example. Using the example of Team A and Team B and assuming independence, we will have the following:

Function 6																
A→; B↓	(-∞,3]	(3,6]	(6,9]	(9,12]	(12,15]	(15,18]	(18,21]	(21,24]	(24,27]	(27,30]	(30,33]	(33,36]	(36,39]	(39,42]	(42,45]	(45,∞)
(-∞,3]	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.11%	0.19%	0.22%	0.19%	0.11%	0.05%	0.01%	0.00%
(3,6]	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.05%	0.13%	0.21%	0.25%	0.21%	0.13%	0.05%	0.02%	0.00%
(6,9]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.10%	0.23%	0.39%	0.47%	0.39%	0.23%	0.10%	0.03%	0.01%
(9,12]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.16%	0.39%	0.66%	0.79%	0.66%	0.39%	0.16%	0.05%	0.01%
(12,15]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.07%	0.25%	0.60%	1.02%	1.22%	1.02%	0.60%	0.25%	0.07%	0.02%
(15,18]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.10%	0.36%	0.85%	1.44%	1.71%	1.44%	0.85%	0.36%	0.10%	0.03%
(18,21]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.13%	0.46%	1.10%	1.85%	2.20%	1.85%	1.10%	0.46%	0.13%	0.03%
(21,24]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.16%	0.54%	1.29%	2.18%	2.60%	2.18%	1.29%	0.54%	0.16%	0.04%
(24,27]	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.17%	0.58%	1.39%	2.35%	2.80%	2.35%	1.39%	0.58%	0.17%	0.04%
(27,30]	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.17%	0.57%	1.37%	2.31%	2.76%	2.31%	1.37%	0.57%	0.17%	0.04%
(30,33]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.15%	0.52%	1.23%	2.09%	2.48%	2.09%	1.23%	0.52%	0.15%	0.04%
(33,36]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.12%	0.42%	1.02%	1.72%	2.05%	1.72%	1.02%	0.42%	0.12%	0.03%
(36,39]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.09%	0.32%	0.77%	1.30%	1.54%	1.30%	0.77%	0.32%	0.09%	0.02%
(39,42]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.06%	0.22%	0.53%	0.89%	1.06%	0.89%	0.53%	0.22%	0.06%	0.02%
(42,45]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.14%	0.33%	0.56%	0.67%	0.56%	0.33%	0.14%	0.04%	0.01%
(45,∞)	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.16%	0.38%	0.64%	0.76%	0.64%	0.38%	0.16%	0.05%	0.01%

Figure 59: The Set-Up of Function 6

Function 6																
A→; B↓	(-∞,3]	(3,6]	(6,9]	(9,12]	(12,15]	(15,18]	(18,21]	(21,24]	(24,27]	(27,30]	(30,33]	(33,36]	(36,39]	(39,42]	(42,45]	(45,∞)
(-∞,3]	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.11%	0.19%	0.22%	0.19%	0.11%	0.05%	0.01%	0.00%
(3,6]	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.05%	0.13%	0.21%	0.25%	0.21%	0.13%	0.05%	0.02%	0.00%
(6,9]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.10%	0.23%	0.39%	0.47%	0.39%	0.23%	0.10%	0.03%	0.01%
(9,12]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.16%	0.39%	0.66%	0.79%	0.66%	0.39%	0.16%	0.05%	0.01%
(12,15]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.07%	0.25%	0.60%	1.02%	1.22%	1.02%	0.60%	0.25%	0.07%	0.02%
(15,18]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.10%	0.36%	0.85%	1.44%	1.71%	1.44%	0.85%	0.36%	0.10%	0.03%
(18,21]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.13%	0.46%	1.10%	1.85%	2.20%	1.85%	1.10%	0.46%	0.13%	0.03%
(21,24]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.16%	0.54%	1.29%	2.18%	2.60%	2.18%	1.29%	0.54%	0.16%	0.04%
(24,27]	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.17%	0.58%	1.39%	2.35%	2.80%	2.35%	1.39%	0.58%	0.17%	0.04%
(27,30]	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.17%	0.57%	1.37%	2.31%	2.76%	2.31%	1.37%	0.57%	0.17%	0.04%
(30,33]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.15%	0.52%	1.23%	2.09%	2.48%	2.09%	1.23%	0.52%	0.15%	0.04%
(33,36]	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.12%	0.42%	1.02%	1.72%	2.05%	1.72%	1.02%	0.42%	0.12%	0.03%
(36,39]	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.09%	0.32%	0.77%	1.30%	1.54%	1.30%	0.77%	0.32%	0.09%	0.02%
(39,42]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.06%	0.22%	0.53%	0.89%	1.06%	0.89%	0.53%	0.22%	0.06%	0.02%
(42,45]	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.14%	0.33%	0.56%	0.67%	0.56%	0.33%	0.14%	0.04%	0.01%
(45,∞)	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.16%	0.38%	0.64%	0.76%	0.64%	0.38%	0.16%	0.05%	0.01%
P(A>B)=	62.25%															

Figure 60: The Output of Function 6 - The Probability A beats B

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Function 7: Determine Seeding Probabilities using Teams' Win/Loss Record Distributions

Overview:

Use truncation and areas to find seeding probabilities

Purpose:

The purpose of this function is to allocate and distribute the number of seeds to each division. This function inputs all teams' win/loss record distributions and output each division's "chance" of entering the playoffs that in total sum up to the number of seeds. The idea is that teams within each division with more wins than losses will have higher chance of entering the playoffs, and thus attaining a seed. To distribute the number of seeds appropriately to each division, two rules need to be followed:

In general, each division has probability of at least 1 (in the next example, it would be 2) of sending one team to the playoff.

For a given division A, the difference (Total number of seeds – divisions' seed probabilities other than division A) should not be less than 1.

Requirements:

I. All teams' win/loss distributions stacked on top of one another. The area under the highest curve should be equal to the number of teams in a league.

Example:

Suppose there is a twelve team league with three divisions and four teams going to the playoffs. The twelve teams' "total season wins" distributions are assumed to be as follows:

Groups:	[0,1,2,3][4,5,6,7][8,9,10,11]																
Seeds	4																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	sum
Prob Dist0	0%	0%	0%	0%	1%	2%	7%	14%	22%	24%	18%	9%	3%	0%	0%	0%	0%	1
Prob Dist1	0%	0%	0%	1%	2%	7%	14%	21%	23%	18%	9%	3%	1%	0%	0%	0%	0%	1
Prob Dist2	0%	0%	0%	0%	2%	6%	13%	20%	23%	19%	11%	4%	1%	0%	0%	0%	0%	1
Prob Dist3	0%	0%	0%	1%	2%	6%	13%	20%	23%	19%	11%	4%	1%	0%	0%	0%	0%	1
Prob Dist4	0%	0%	1%	4%	11%	18%	22%	20%	14%	7%	2%	1%	0%	0%	0%	0%	0%	1
Prob Dist5	0%	0%	1%	4%	11%	18%	23%	20%	13%	6%	2%	0%	0%	0%	0%	0%	0%	1
Prob Dist6	0%	0%	2%	7%	15%	22%	23%	17%	9%	4%	1%	0%	0%	0%	0%	0%	0%	1
Prob Dist6	0%	0%	1%	4%	9%	17%	22%	21%	15%	8%	3%	1%	0%	0%	0%	0%	0%	1
Prob Dist8	0%	1%	4%	11%	19%	23%	20%	12%	6%	2%	0%	0%	0%	0%	0%	0%	0%	1
Prob Dist9	0%	1%	3%	8%	16%	22%	22%	16%	9%	3%	1%	0%	0%	0%	0%	0%	0%	1
Prob Dist10	0%	1%	3%	9%	17%	22%	21%	15%	8%	3%	1%	0%	0%	0%	0%	0%	0%	1
Prob Dist11	0%	1%	3%	9%	17%	22%	22%	15%	8%	3%	1%	0%	0%	0%	0%	0%	0%	1

Division A: Teams 0, 1, 2, 3

Division B: Teams 4, 5, 6, 7

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Division C: Teams 8, 9, 10, 11

The numbers 0-16 in the first row represent the bins; 0 being the 0th bin, and 16 being the 16th bin. The win/loss record ascends from left to right; so on the leftmost is the lowest win/loss record and on the rightmost is the highest win/loss record. The “sum” column indicates that each team’s distribution adds up to 1. The total of the “sum” column should be equal to the number of teams in a league.

In this example, the total number of seeds is equal to 4. Each division sends at least one team to the playoffs and one division will send a second team as a wild-card.

Here is a graphical representation of how the distributions of those 12 teams stack on top of another:

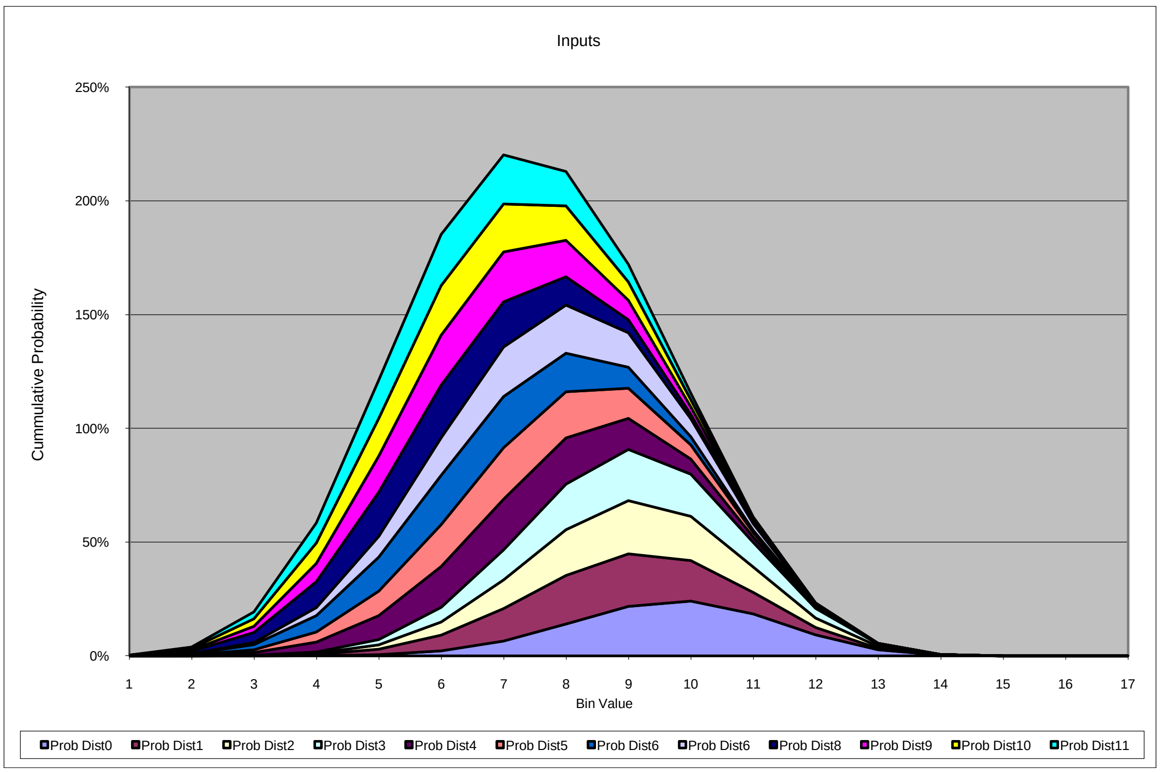


Figure 61: A 12 Team Win/Loss Distribution Stack

Based on the assumptions that:

I. The probability of a certain team landing on the rightmost edges on the graph with high win/loss record should correspond to its chance of being the top seeds.
II. And the sum of probability of each team being seed 1 is one (or there is going to be a seed 1 in these teams).
III. A division has to send a minimum of one team to the playoffs.

Here is the output using Function 7:

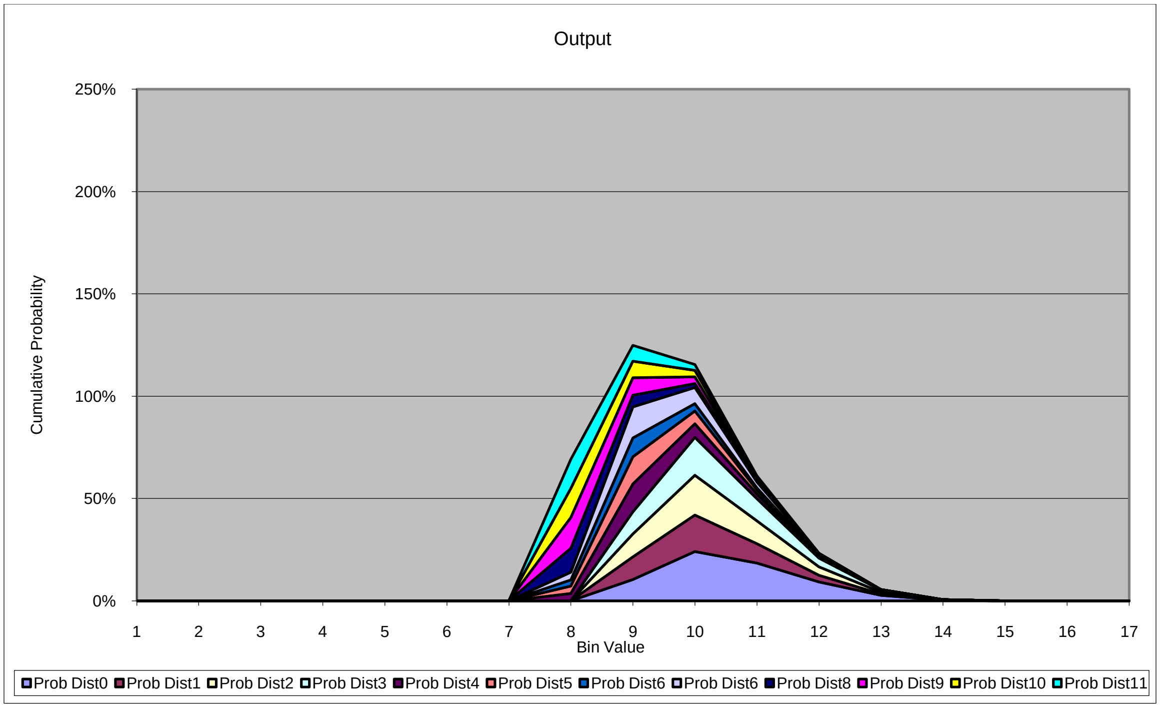


Figure 62: Output of Function 7

Mathematical Background:

At this stage in the season calculator each team has its own win/loss record curve, which represents the probabilities (on the y-axis) that a team will end the season with a specific record (on the x-axis). The area under the curve for each team is equal to one, since the team must finish the season with one of the designated records. All of the win/loss records for the teams in the league are then layered on top of each other, in no particular order. The sum of the area under all the curves is equivalent to the number of teams in the league. Then, starting from the rightmost edge of the graph, move from right to left until the cumulative area under all of the curves is equal to one. This section of the graph represents "seed 1". The portion of that section of the graph that belongs to each team is equivalent to their percentage chance of being "seed 1".

In order to identify the probability that each team is "seed 2" first draw a vertical line that marks the boundary for "seed 1". Then, move from the right to the left until the area under the curves, after the "seed 1" boundary line is equal to one. This section represents "seed 2". The portion of this section of the graph that belongs to each team is equivalent to their percentage chance of being "seed 2". Then, repeat this process to find all of the remaining seeds.

A team's chance of making it to the playoffs is the sum of their percentage chance of being the first however many seeds go to the playoffs, which is league specific, and initially entered by the user. For example, if Team X belongs to a league where the first four seeds go to the playoffs, and they have a

10% chance of being seed 1, a 12% chance of being seed 2, a 13% chance of being seed 3, and a 15% chance of being seed 4, then they have a 50% chance of making it to the playoffs.

In order to account for divisions within a league only a slight change needs to be made. The total cumulative probability of a division being any seed cannot exceed the number of teams in that division. To show how this will be done, let us first start with an example. In League Y, the first six seeded teams will go to the playoffs, and there are three divisions. The percentage chance of any team going to the playoffs is equal to the sum of the percentage chance of being seeds 1-6. Using our current method it is possible, since we are currently not accounting for divisions, that not all divisions will send a team to the playoffs. Therefore, it is necessary, when moving from right to left, to stop once one division no longer has a probability of at least one of going to the playoffs.

Each division must have a probability of at least one of sending a team to the playoffs, and in order to ensure that this is true, it is necessary to truncate the other division's chances once a division no longer has a chance of sending at least one team to the playoffs. Once the other division's curves have been truncated, then move from right to left until the area under the curves of only the teams in the remaining division are equivalent to one.

Mathematical Formulas:

12 team league, 3 divisions, 4 teams in each division, Seeds 1-6 go to playoffs:

Division 1: Teams 1, 2, 3, 4; Division 2: Teams 5, 6, 7, 8; Division 3: Teams 9, 10, 11, 12

Seed 1: $\int_{S1}^{S0} \sum_{i=1}^{12} T_i \, dx = 1$, where "S0" represents the rightmost point on the graph created by the aggregate of Teams 1-12. This equation is used to find the value of S1.

Once you have calculated S1, you can determine the probability that a team from a certain division is seed 1:

Division 1, Seed 1: $\int_{S1}^{S0} \sum_{i=1}^{4} T_i \, dx = \Pr\text{ (Division 1 has a Seed 1 team)}$

Division 2, Seed 1: $\int_{S1}^{S0} \sum_{i=5}^{8} T_i \, dx = \Pr\text{ (Division 2 has a Seed 1 team)}$

Division 3, Seed 1: $\int_{S1}^{S0} \sum_{i=9}^{12} T_i \, dx = \Pr\text{ (Division 3 has a Seed 1 team)}$

You can also use this value of S1 to determine the probability that a certain team is seed 1:

Team 1, Seed 1: $\int_{S1}^{S0} T_1 \, dx = \Pr\text{ (Team 1 is Seed 1)}$

Team 2, Seed 1: $\int_{S1}^{S0} T_2 \, dx = \Pr\text{ (Team 2 is Seed 1)}$

Team 3, Seed 1: $\int_{S1}^{S0} T_3 \, dx = \Pr\text{ (Team 3 is Seed 1)}$ etc.

Then, you need to identify the new boundary line for seed 2:

Seed 2: $\int_{S2}^{S1} \sum_{i=1}^{12} T_i \, dx = 1$

Once you have calculated the value of S2, you calculate the probabilities for both divisions and teams in the same way as defined above.

For Division 1: $\sum_{i=1}^{6} \int_{Si}^{S(i-1)} \sum_{j=1}^{4} T_j \, dx \geq 1$

For Division 2: $\sum_{i=1}^{6} \int_{Si}^{S(i-1)} \sum_{j=5}^{8} T_j \, dx \geq 1$

For Division 3: $\sum_{i=1}^{6} \int_{Si}^{S(i-1)} \sum_{j=8}^{12} T_j \, dx \geq 1$

The chance that team i goes to the playoffs is:

$$\sum_{i=1}^{6} \int_{Si}^{S(i-1)} T_i \, dx$$

Function 8: Develop a Win/Loss Probability Tree

Overview:

Find the probability tree of win/loss record using probability of wins for each week

Purpose:

The purpose of this function is to develop a win/loss probability tree which correlates to a win/loss distribution for a specific team, given the team's chance of winning against his opponent in weekly matchups. The input of this function is N probabilities, (N being the number of weeks a team has a matchup). The output of this function is a distribution of win/loss records and their associated probabilities.

Requirements:

I. The win/loss probability output of Function 6 for each week.

Example:

Suppose Team A plays Team B in Week 1 and Team A plays Team C in Week 2. The probability of Team A win their first matchup is 75%, the probability of Team A wins their second matchup is 45%. Assuming the two matchups are independent, we can develop a probability tree that looks like the following:

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

Team A	Week 1	Week 2
p =	0.75	0.45
q =	0.25	0.55
		2--0
	1--0	33.75%
0--0	75.00%	1--1
100.00%	0--1	52.50%
	25.00%	0--2
		13.75%

Figure 63: Team A Probabilities of Winning in Week 1 and 2

For Team A this gives you all of the possible win/loss records at the end of the two weeks and their associated probabilities. In this example, after two weeks, Team A will have a 33.75% chance of having 2 wins, a 52.5% chance of having 1 win and 1 loss, and a 13.75% chance of having 2 losses.

The idea can be extended to multiple weeks with probability of winning that week being given. For more information, please see Phase 5.

Function 9: Develop Playoff Win/Loss Distribution for Each Team (Using Probability Trees)

Overview:

Create the probability tree of win/loss record during playoffs using probability of wins for each week

Purpose:

The purpose of this function is to develop a probability distribution of the possible win/loss records for each team entering the playoffs. It uses the output from Function 7, the probability of a team obtaining certain seeds and calculates the team beats its opponents during matchups in the playoffs.

Requirements:

I. Seeding probability from Function 7
II. Win/Loss Probability Tree from Function 8

Example:

Please refer to Phase 5.

Function 10: Calculate the Expectation of a Random Variable X.

Overview:

Find the expectation of random variable X

Purpose:

The purpose of this function is to calculate an expectation of a random variable X.

Requirements:

I. A probability distribution

Mathematical formula:

$$E(X) = \int_{-\infty}^{\infty} x * f(x)dx$$

Or,

$$E(X) = \sum_{X=x} x * P(X = x)$$

How to Calculate STDpp and STDpa

The fantasy point projection of each player will change from week to week. Suppose you are in Week 2 and your player's projection for Week 15 is 20 fantasy points. In Week 14, your player's projection for Week 15 is 15 fantasy points. The difference of 5 fantasy points (20 minus 15) represents the variation over the span of 12 weeks. This variation is used to calculate the 1-week STDpp for a specific position.

There are about 1,400 football players in the NFL and about 200 to 300 players in each position. Due to large computations involved in computing the 1-week STDpp for a specific position, only the top 30 players from each position will be used for the calculation of 1-week STDpp for a given position. They will be ranked in descending order by fantasy points for the week you choose, (in the following example, it will be Week 2) and the same players' fantasy points will be recorded for that week. To develop STDpp, you will also need projections for Week 15 developed at a time other than during Week 2 – for this example, we'll use Week 14. So we have two projections for Week 15, one developed from AccuScore data as of Week 2, and one developed twelve weeks later, as of Week 14. The two weeks you choose can be any weeks other than Week 15. The difference between the two weeks' projections will be used to develop STDpp, a measure of how projections vary over time.

Suppose I have received the fantasy point projections for Week 15 from top 30 quarterback scorers in Week 2 from AccuScore, and I receive new projections for all 30 players in Week 14, as shown below:

Player ID	Week 2's Week 15	Week 14's Week 15
27	25	22
17	25	23
12	25	24
19	24	23
6	24	22
10	23	18
8	23	17
4	23	18
28	22	16
25	22	15
15	22	20
14	22	22
11	22	17
7	22	18
29	21	15
20	20	20
5	20	20
26	19	17
24	19	15
21	19	22
16	19	25
13	19	20
30	18	21
18	18	15
3	18	15
23	17	18
1	17	18
22	15	19
9	15	18
2	15	25

Figure 64: STDpp Calculations

The first column indicates Players' ID, or could be their actual name; the second column is their Week 2's projections of the amount of fantasy points they will score in Week 15; and the third column is their Week 14's projections of the amount of fantasy points they will score in Week 15.

Below shows how to calculate 1-week STDpp for quarterbacks in this example:

Player ID	Week 2's Week 15	Week 14's Week 15	Δ	Δ^2
27	25	22	3	9
17	25	23	2	4
12	25	24	1	1
19	24	23	1	1
6	24	22	2	4
10	23	18	5	25
8	23	17	6	36
4	23	18	5	25
28	22	16	6	36
25	22	15	7	49
15	22	20	2	4
14	22	22	0	0
11	22	17	5	25
7	22	18	4	16
29	21	15	6	36
20	20	20	0	0
5	20	20	0	0
26	19	17	2	4
24	19	15	4	16
21	19	22	-3	9
16	19	25	-6	36
13	19	20	-1	1
30	18	21	-3	9
18	18	15	3	9
3	18	15	3	9
23	17	18	-1	1
1	17	18	-1	1
22	15	19	-4	16
9	15	18	-3	9
2	15	25	-10	100
		sum:	35	491
1-week STDpp=		sqrt(491/(14-2))		
1-week STDpp=		6.396613687		

Figure 65: STDpp Calculations 2

In addition to determining a measure of possible variance in projections using STDpp as described above, we are also interested in the accuracy of projections for a given week. For this, we develop a measure called STDpa, which is an estimate of how much a particular week's projections might vary from reality. STDpa's calculations would be very similar, except that we will use the projection at the beginning of Week 15 for Week 15's projection, and compare this to the actual result in Week 15. As shown below:

Player ID	Week 15's Week 15	Actual Result Week 15	Δ	Δ^2
27	25	22	3	9
17	25	23	2	4
12	25	24	1	1
19	24	23	1	1
6	24	22	2	4
10	23	18	5	25
8	23	17	6	36
4	23	18	5	25
28	22	16	6	36
25	22	15	7	49
15	22	20	2	4
14	22	22	0	0
11	22	17	5	25
7	22	18	4	16
29	21	15	6	36
20	20	20	0	0
5	20	20	0	0
26	19	17	2	4
24	19	15	4	16
21	19	22	-3	9
16	19	25	-6	36
13	19	20	-1	1
30	18	21	-3	9
18	18	15	3	9
3	18	15	3	9
23	17	18	-1	1
1	17	18	-1	1
22	15	19	-4	16
9	15	18	-3	9
2	15	25	-10	100
		sum:	35	491
STDpa=		sqrt(491/30)		
STDpa=		4.045573713		

Figure 66: STDpa Calculations

Summary:

STDpp: measures the variation of projection to projection
STDpa: measures the variation of projection to actual result

Recommendations

Through the course of our numerous discussions with the CEO of Advanced Sports Logic, we identified a few areas of improvements which could not be addressed within this project, but may warrant further investigation. We recommend that Advance Sports Logic examine the possibility of using a non-normal distribution for their player projections, obtaining projections from other sources than AccuScore, treating wins and losses as dependent rather than independent events, improving calculations for STDpp and STDpa and setting up partnerships with Fantasy Football companies.

I. Using Non-Normal Fantasy Point Distributions

Currently, the season calculator receives a projection from AccuScore which represents the number of fantasy points a football player is projected to score. This number is then used by the season calculator as the mean of a normal distribution curve, and every player within a certain position is assigned the same standard deviation. For example, all quarterbacks have the same standard deviation, which is different from the standard deviation that all wide receivers have.

We believe that using the same standard deviation for all types of players within a certain position is potentially flawed, and propose instead that the season calculator break up players into three categories: elite, average and bench. Elite players would be the ones with the highest fantasy point projected mean, but the area under their probability curve would be skewed to the left. It is more likely that if their actual fantasy points scored differs greatly from their projected mean, it will differ in a negative way, and they will score lower than their projected mean. Since they are already very strong players this means that they play the majority of the games for their team and are more likely to suffer an injury as well. Average players would still have equal chances of scoring higher or lower than their projected mean. Bench players would be the ones with the lowest fantasy point projected mean, but the area under their probability curve would be skewed to the right. Since these players receive less playing time for their actual NFL teams, it is more likely that if their actual fantasy points scored differs greatly from their projected mean, it will differ in a positive way, and they will score higher than their projected mean.

II. Obtaining Projections from Other Sources

While AccuScore thus far has provided valuable raw data for the season calculator, we believe it would be in the best interests of Advanced Sports Logic to supplement the projections receive with projections from other sources, such as NFL.com or CBS. As the 2009-10 NFL season drew to a close the CEO of Advanced Sports Logic told us that AccuScore appeared to consistently under-project Tom Brady and Peyton Manning, and consistently over-project the entire Minnesota Vikings starting roster. Since AccuScore does provide a large amount of valuable data we feel that Advanced Sports Logic should still utilize them as their main source of raw data, however to address this problem of under- and over-projections, we recommend that ASL supplement their data with that of other NFL projection companies. One way to get started on this recommendation would be to do a study of how well AccuScore's projections have lined up with reality the past few years (or for whatever time period data can be gathered). On an ongoing basis, it would be good to capture AccuScore's projections, so that in the future, a study of their accuracy will be easier to undertake.

III. Treating Wins/Losses as Dependent Events

Currently the season calculator treats the probabilities each team has of winning or losing a matchup as independent events. However, a team cannot have a completely winning record unless all of the teams they face in matchups lose their matchups. Therefore, whether a team wins or loses is not independent of another team's performance. The assumption of independence may or may not have a major impact on the projections that are being done currently, but this ought to be investigated.

At this time we cannot provide any specific recommendations to address this issue, and we do recognize that any solution to this problem may be quite complicated and computationally intensive. We do however recommend that at some point Advanced Sports Logic examine the possibility of treating wins and losses as dependent events.

IV. Improve STDpp and STDpa calculations

Another possibility of calculating STDpp is that if we assume that the top 30 players at each position will have the same distribution and we denote this as random variable X, with its mean and variance. Then their fantasy point projection of a particular week will be random samples or "observations" from this random variable X. If we use the formula:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$E\overline{X} = \mu$$

$$ES^2 = \sigma^2$$

When n is large, $\overline{X}$ will be approximately the true mean, and S^2 will be approximately the true standard deviation, both are unbiased estimators.

Using the same example above, we can compute the $\overline{X}$ and S^2 as follows:

Player ID	Week 2's Week 15	Average of Column B	Difference of B and C	Square of Column D	Sum of Column E	1/(n-1)= 0.0344828	Multiply F and G
27	25	20.43	4.57	20.85	263.37	0.0344828	9.0816092
17	25	20.43	4.57	20.85	263.37	0.0344828	9.0816092
12	25	20.43	4.57	20.85	263.37	0.0344828	9.0816092
19	24	20.43	3.57	12.72	263.37	0.0344828	9.0816092
6	24	20.43	3.57	12.72	263.37	0.0344828	9.0816092
10	23	20.43	2.57	6.59	263.37	0.0344828	9.0816092
8	23	20.43	2.57	6.59	263.37	0.0344828	9.0816092
4	23	20.43	2.57	6.59	263.37	0.0344828	9.0816092
28	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
25	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
15	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
14	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
11	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
7	22	20.43	1.57	2.45	263.37	0.0344828	9.0816092
29	21	20.43	0.57	0.32	263.37	0.0344828	9.0816092
20	20	20.43	-0.43	0.19	263.37	0.0344828	9.0816092
5	20	20.43	-0.43	0.19	263.37	0.0344828	9.0816092
26	19	20.43	-1.43	2.05	263.37	0.0344828	9.0816092
24	19	20.43	-1.43	2.05	263.37	0.0344828	9.0816092
21	19	20.43	-1.43	2.05	263.37	0.0344828	9.0816092
16	19	20.43	-1.43	2.05	263.37	0.0344828	9.0816092
13	19	20.43	-1.43	2.05	263.37	0.0344828	9.0816092
30	18	20.43	-2.43	5.92	263.37	0.0344828	9.0816092
18	18	20.43	-2.43	5.92	263.37	0.0344828	9.0816092
3	18	20.43	-2.43	5.92	263.37	0.0344828	9.0816092
23	17	20.43	-3.43	11.79	263.37	0.0344828	9.0816092
1	17	20.43	-3.43	11.79	263.37	0.0344828	9.0816092
22	15	20.43	-5.43	29.52	263.37	0.0344828	9.0816092
9	15	20.43	-5.43	29.52	263.37	0.0344828	9.0816092
2	15	20.43	-5.43	29.52	263.37	0.0344828	9.0816092
sample mean =		20.43			sample variance =		9.0816092

Figure 67: Propose STDpp & STDpa Calculations

The sample mean and sample variance may be depend on the specific week, specific players.

V. Partnerships with Fantasy Football Companies

Currently, Advanced Sports Logic is in the process of developing relationships with CBSSports and MyFantasyLeague.com to provide this software to their users. This allows CBSSports and MyFantasyLeague.com users to directly import their data from the website, rather than manually enter their information. This will save the team manager a lot of time, and will ensure that the season calculator has the most up-to-date, correct information.

Team managers who do not use CBS or MyFantasySports.com have to manually enter all of the league-specific information into the season calculator, as well as all of the draft picks as they occur. They also need to frequently update the season calculator whenever another team manager alters their roster. This process can be tedious and time consuming, and may be a turn off for potential customers.

We strongly recommend that Advanced Sports Logic continue to develop partnerships with all major fantasy football providers, as well as seek out methods by which to advertise their product on the providers' websites. This will help ASL attract more customers, which will increase brand recognition as well as profitability, and will prove to be extremely valuable to the success of the company.

Conclusion

The purpose of our project was to analyze the methods and functions that were in place, and propose more accurate, precise methods for implementation in the future. We divided the operations performed by the season calculator into phases, and organized them into this flow chart. Within these phases, different functions are used. As we examined each phase and function, we attempted to identify what would be the most mathematically accurate way to achieve the desired result of that phase or function. Some of these approaches would require very large numbers of calculations, making them impractical for the software. Since the season calculator needs to be able to run during a draft and aid in the selection of players, our methods needed to be modified, to be usable given the constraints of both the time limit during the draft, as well as the constraints imposed by the client's computer on which the calculations are performed.

Over the course of the project, we have come up with a few recommendations that the next MQP team, as well as Advanced Sports Logic may be inclined to investigate further. These are discussed in the preceding section, and could well form the basis for another MQP during the 2011-12 academic year.

In conclusion, we greatly enjoyed the opportunity to work with Advanced Sports Logic. We helped them describe their current methods in mathematical terms and helped propose new, more accurate methods which could be implemented in the software. We investigated and identified the most precise way to calculate the probability that a fantasy football team wins its playoffs, and modified them to operate under our given constraints of both time and computing power. We also provided recommendations for Advanced Sports Logic, as well as any potential future MQP team for what they might choose to investigate and improve in the future.

We would like to thank Leonard LaPadula, CEO of the Advanced Sports Logic for providing us with this great opportunity to help on his project. We would also like to thank Professor Jon Abraham for his valuable guidance throughout the project and his insightful comments.

Bibliography

Accuscore Advisor. (n.d.). Retrieved November 2010, from FAQ: http://accuscore.com/faq

Advanced Sports Logic. (2010). Retrieved November 2010, from The Machine: http://www.advancedsportslogic.com/

No formulas or definitions were extracted directly from the following sources; however they were used to gain a further understanding of the ideas that were implemented:

Casella, George, and Roger L. Berger. *Statistical inference*. 2nd ed. Pacific Grove, CA: Duxbury Pr, 2002. Print.

Abramowitz, Milton, and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover Publications, 1965. Print.

Appendices

Appendix A: Fantasy Football

American Fantasy Football is a mainly virtual competition between teams consisting of actual National Football League (NFL) players which are run by mangers (the users). In general, each manager drafts players from different teams within the NFL for their own fantasy football team roster before the actual NFL season begins. Then, based on the league specific rules, teams are awarded a certain number of fantasy points based on the performance of their drafted players in the NFL matchups that week.

A fantasy football team consists of one quarterback (QB), three wide receivers (WR), two running backs (RB), one tight end (TE), one placekicker (K), one NFL team defense/special teams (DEF / DST), and six bench players (BN) which can fill any of the aforementioned positions. There are some leagues that allow players to draft individual players for all defensive positions, and others that allow managers to draft players for every playable position. Ideally managers want starting players that will score the most points, and backup players to fill in open roster spots during a bye week or when a starter is injured.

In a “head-to-head” league, fantasy football teams are matched up in a schedule similar to the NFL. The team that has the most fantasy points from their starting roster at the end of the football week earns a win for their manager. In this sense, a low fantasy point scoring team earning only 73 fantasy points in one week could win their matchup if their opponent’s players earned less than 73 fantasy points in total. Similarly, a high scoring team that earned 130 fantasy points one week could still lose their matchup if their opponent’s players earned 142 fantasy points. This makes strategy significantly more important than points in certain situations. For example, a manager may choose to play a WR to compliment their opponent’s QB in hopes that the WR may catch a touchdown (TD) pass, thus boosting each teams’ points relatively equally (depending on the scoring rules of the league). In “head-to-head” leagues the fantasy points are important, but finding a strategy to earn more fantasy points than one’s opponent is crucial. The teams with the most amount of wins by the end of the season move on to the play-offs. In “total points” leagues, teams with the greatest number of fantasy points are the ones that move on to the play-offs. There are no matchups per week, and teams are ranked by their overall season fantasy points earned to date. Fantasy points are essential to success in these leagues.

Throughout the season, managers can add, drop, and trade players as well. The trading process is self-explanatory, although some leagues allow other managers not involved in the trade to veto transactions that they may find unfair. The trade is then eliminated after a certain number of vetoes are expressed. Trades can also consist of uneven numbered trades, where one manager trades two players for one player from another team’s roster. Depending on the league, when adding players to their roster, managers have the option of choosing either a free agent, or picking up a player on waivers. Free agents are players who can be added to a team at any time. Waivers take a certain amount of time to process and add the player. During this time, another manager could attempt to add the player as well. When the waiver period is up, the manager with the lower waiver number will end up with the player. This allows for slight competition in picking up players. The waiver number is determined by rank and fantasy point totals; this determination varies greatly from league to league. Managers usually

Advanced Sport Logic, Inc. & Worcester Polytechnic Institute Confidential

drop unwanted players which makes them newly available free agents. In some leagues however, managers cannot drop certain players from their team. There are also leagues that allow managers to place their own players on waivers, thus effectively dropping them, yet making them unavailable to be picked up by other players until the waiver period expires. There are many different nuances in fantasy football, which allows for many different leagues to be formed. This versatility draws people to the game of fantasy football; people can find a league that best suits them and their hopes of winning it all.

Appendix B: AccuScore

AccuScore is a fantasy-sports projection and investment company run out of Los Angeles, California. It is considered to be one of the largest sports investment support websites. AccuScore itself states that, “AccuScore.com does not take bets, but rather 100% legally provides information [and] support for both sports investment as well as fantasy/rotisserie leagues.” Accuscore provides player projections for all players within the NFL. These projections are done by 10,000 game simulations using AccuScore’s scoring rules, and thusly represent the mean fantasy points a player is projected to score. In addition to the weekly projections, AccuScore also provides full season statistical projections for each player. These comprehensive weekly player projections are used by Advanced Sports Logic Inc. to create fantasy point distributions for each player (Accuscore Advisor).

Appendix C: Advanced Sports Logic

Advanced Sports Logic Inc. is a service based entrepreneurial company which offers its customers software to guide the user in improving their likelihood of winning their fantasy football league. The software which is known as "The Machine", provides recommendations for all possible transactions within an average fantasy football league, such as draft picks, starting lineups, and trade proposals. ASL receives weekly statistical player projections from AccuScore, and combines them with historical projection variance and accuracy data along with player injury data to convert them into fantasy point probability distributions. They then develop an aggregate team fantasy point distribution, and proceed to simulate the remainder of the season. Instead of predicting whether a matchup is won or lost, or predicting which seed the customer's team is likely to obtain, ASL provides their users with probabilities for each situation. ASL differs from their competitors in this aspect; their goal is to use rigorous mathematical methods to determine a team's chance to win the championship (Advanced Sports Logic, 2010).

Appendix D: Normal Approximation

According to Abramowitz and Stegun's Handbook of Mathematical Functions 7.1.26 and the definition of the error function, the standard normal cumulative distribution function Φ(x) can be approximated as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

Where

$$\operatorname{erf}(y) = 1 - (at + bt^2 + ct^3 + dt^4 + et^5) \times \exp(-y^2) + \operatorname{error}(y),$$

$$t = \frac{1}{1 + py}$$

$$|\operatorname{error}(y)| \leq 1.5 \times 10^{-7}$$

$$p = .3275911, a = .254829592, b = -.284496736,$$

$$c = 1.421413741, d = -1.453152027, e = 1.061405429$$

x=	1.26	INPUT
x/2½=	0.890954544	
a=	0.254829592	
b=	-0.284496736	
c=	1.421413741	
d=	-1.453152027	
e=	1.061405429	
p=	0.3275911	
1/(1+px)=	0.774072426	
rf (x/2½)=	0.79	
Φ(x)=	0.89616525	OUTPUT
≈	0.8961653	
T(x,0,1,1)=		
	0.89616532	Excel Approximation
≈	0.8961653	
The approximation is very close up to the 7th decimal place		
The largest error is bounded by 1.5*10^(-7)		

Figure 68: Approximation of Normal vs. Excel

This spreadsheet uses an estimating procedure to develop cumulative probabilities under a Normal Distribution with mean 0 and standard deviation 1. It achieves this by using a well-known approximation for the error function (ERF) and using a simple algebraic manipulation to relate ERF and

the Normal Distribution. The ERF approximation, presented in the Handbook of Mathematical Functions (edited by Abramowitz and Stegun), is considered to be very accurate. Because it was published by the National Standards Board, there is no copyright on the book, and it is freely in the public domain. A link to the specific formula implemented here is given below.
http://people.math.sfu.ca/~cbm/aands/page_299.htm

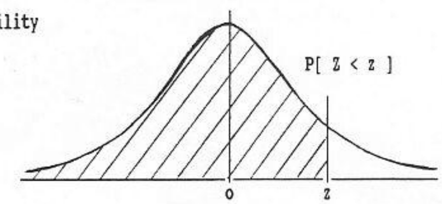
Here is a Standard Normal Cumulative Distribution Function Φ(x) Table:

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z
i.e.

$$P[\ Z < z\ ] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Figure 69: CDF of Standard Normal Table