Sports Analytics Final Project

Overview of Problem

For our final project, our question is: **When teams are in "field goal range", should they settle for a field goal attempt on 4th down rather than going for a touchdown?**

To answer this question, we started by defining the following two fourth down strategies:

- Strategy A: Always kick a field goal if you are on or beyond your opponent's 40 yard-line and it's 4th down
- Strategy B: Always go for-it if you are on or beyond your opponent's 40 on 4th down

Ultimately, we will use this to calculate the expected points of each of these strategies for a range of field position-yards to first down combinations, and a break-even probability of success. The break-even probability of success is, at minimum, how confident a coach must be in his team's ability to convert on 4th down for it to make sense to go for it rather than kicking a field goal.

Methodology

Our group chose to use R as our programming language for the purposes of this project. We selected R due to our previous experience using it.

*Fitting Distributions:* Using ggplot and the fitdistR packages in R, we were able to create mixture models to fit distributions to various game-time situations. After conducting preliminary EDA, we determined that it was appropriate to develop a model for 1st and 2nd down scenarios, and separate models for both 3rd and 4th down. We then created bins for field position and yards to first down. Next, we used the data to determine how often teams run vs. pass in each of these situations. From there, we were able to visualize the distribution of yards gained in each of these scenarios, and fit an empirical distribution.

*Building Simulation*: After we fit distributions for each of our scenarios, our next task was to create a function that would simulate an actual game until the point where a team scores, which could then be used to calculate the expected points for a particular starting state (effectively employing Markov Chains). Our function included the following parameters: **field position, yards to first down, down, 4th down strategy, starting possession.**

*We used the following formulas to calculate expected points:*

**E[Kicking a Field Goal]** = $P_{success}$(3 + E[Opp 1st & 10 at own 25]) + (1-$P_{success}$)(E[Opp 1st & 10 at FP - 7yds])

**E[Going For It]** = $P_{success}$(E[1st & 10 at FP+Yards Gained on last play]) = (1-$P_{success}$)(E[Opp 1st & 10 at FP+Yards Gained on last play])

Under the assumption that the probability of making a field goal is fixed using a logistic regression model, we can set E[Kicking a Field Goal] equal to E[Going for It] and solve for the break-even probability of success, given a specific range for field position and yards to first down. These results can be found below.

Results: Break-Even Probability

A = Expected points of kicking a field goal
B = Expected points after converting on 4th down
C = Expected points of opponent after not converting on 4th down

Note: FP describes the number of yards past a team's own end zone. Thus, when FP = 65, the team is actually on the opponent's 35 yard line.

| FP | YTG | A = E(FG) | B = E[1st & 10 at FP+YTG] | C = E[Opp 1st & 10 at FP] | P(Break-Even) |
|---|---|---|---|---|---|
| 60-68 | 0-1 | -0.0138 | 3.0046 | -1.3496 | 0.3068 |
| 60-68 | 2-3 | -0.0138 | 3.1476 | -1.3496 | 0.2970 |
| 60-68 | 4-5 | -0.0138 | 3.2706 | -1.3496 | 0.2891 |
| 60-68 | 6-7 | -0.0138 | 3.2258 | -1.3496 | 0.2920 |
| 60-68 | 8-10 | -0.0138 | 3.2608 | -1.3496 | 0.2897 |
| 69-77 | 0-1 | 0.7299 | 3.679 | -0.997 | 0.3693 |
| 69-77 | 2-3 | 0.7299 | 3.8222 | -0.997 | 0.3583 |
| 69-77 | 4-5 | 0.7299 | 3.9246 | -0.997 | 0.3509 |
| 69-77 | 6-7 | 0.7299 | 4.008 | -0.997 | 0.3450 |
| 69-77 | 8-10 | 0.7299 | 4.1564 | -0.997 | 0.3351 |
| 78-86 | 0-1 | 1.2910 | 4.2554 | -0.8652 | 0.4211 |
| 78-86 | 2-3 | 1.2910 | 4.3214 | -0.8652 | 0.4157 |
| 78-86 | 4-5 | 1.2910 | 4.4818 | -0.8652 | 0.4033 |
| 78-86 | 6-7 | 1.2910 | 4.7232 | -0.8652 | 0.3858 |
| 78-86 | 8-10 | 1.2910 | 5.125 | -0.8652 | 0.3600 |
| 87-94 | 0-1 | 1.6477 | 5.0576 | -0.539 | 0.3907 |
| 87-94 | 2-3 | 1.6477 | 5.5034 | -0.539 | 0.3619 |
| 87-94 | 4-5 | 1.6477 | 5.8548 | -0.539 | 0.3420 |
| 87-94 | 6-7 | 1.6477 | 6.275 | -0.539 | 0.3209 |
| 87-94 | 8-10 | 1.6477 | 6.6334 | -0.539 | 0.3049 |
| 95-100 | 0-1 | 1.7965 | 6.4444 | -0.5286 | 0.3334 |
| 95-100 | 2-3 | 1.7965 | 6.7724 | -0.5286 | 0.3185 |
| 95-100 | 4-5 | 1.7965 | 6.8826 | -0.5286 | 0.3137 |

A, B and C are calculated using 10,000 iterations of our simulation described above.

Results: General Strategy Based on Empirical Simulation

This table calculates the expected points of kicking a field goal and going-for-it on 4th down for the average team, as determined by our simulation with 10,000 iterations for each scenario. This table can be used by coaches to develop general strategies for play calling as a supplement to the break-even probability of success.

| FP | YTG | EP(FG) | EP(Go for It) | Strategy |
|---|---|---|---|---|
| 60-68 | 0-1 | -0.0138 | 0.9686 | Go for It |
| 60-68 | 2-3 | -0.0138 | 0.7088 | Go for It |
| 60-68 | 4-5 | -0.0138 | 0.3964 | Go for It |
| 60-68 | 6-7 | -0.0138 | 0.0834 | Go for It |
| 60-68 | 8-10 | -0.0138 | -0.2781 | FG |
| 69-77 | 0-1 | 0.7299 | 1.5322 | Go for It |
| 69-77 | 2-3 | 0.7299 | 1.2107 | Go for It |
| 69-77 | 4-5 | 0.7299 | 0.8757 | Go for It |
| 69-77 | 6-7 | 0.7299 | 0.5305 | FG |
| 69-77 | 8-10 | 0.7299 | 0.2517 | FG |
| 78-86 | 0-1 | 1.2910 | 2.2732 | Go for It |
| 78-86 | 2-3 | 1.2910 | 1.2629 | FG |
| 78-86 | 4-5 | 1.2910 | 0.8143 | FG |
| 78-86 | 6-7 | 1.2910 | 0.4319 | FG |
| 78-86 | 8-10 | 1.2910 | 0.0675 | FG |
| 87-94 | 0-1 | 1.6477 | 3.0971 | Go for It |
| 87-94 | 2-3 | 1.6477 | 1.8242 | Go for It |
| 87-94 | 4-5 | 1.6477 | 1.2730 | FG |
| 87-94 | 6-7 | 1.6477 | 0.7216 | FG |
| 87-94 | 8-10 | 1.6477 | 0.2844 | FG |
| 95-100 | 0-1 | 1.7965 | 4.0854 | Go for It |
| 95-100 | 2-3 | 1.7965 | 2.3604 | Go for It |
| 95-100 | 4-5 | 1.7965 | 1.5903 | FG |

Analysis and Conclusions

As we can see above, at a field position ranging from 60 to 68, the break-even probability of success goes down as the yards-to-go increases (with the exception of yards-to-go values of 4 to 5, where we saw a slight decrease). We can see that for the 60 to 68 field position, there is a range of values from approximately 0.29 to 0.31.

When looking at the field position ranging from 69 to 77, we see the same trend. As yards-to-go increases, the break-even probability of success decreases. We see a range of values from approximately 0.34 to 0.37. We can see that while our values are slightly higher than compared to the field position from 60 to 68, they are still relatively low, pointing towards the 0.3 range.

When looking at the next bin of field position values from 78 to 86, we can once again see the same trend. This time, our range is from 0.36 to 0.42. The probability of approximately 0.42 is the highest probability that we have yet witnessed in our analysis.

Our second to last bin is the field position ranging from 87 to 94. Here, once again, we see the same trend. Interestingly, our highest break-even probability of success is 0.39. This is significantly lower than the upper range for the analysis with the field position ranging from 78 to 86. This is interesting because we will see, especially with the next bin for the field position, that our break-even probability of success is going down slightly. Overall, for this grouping, we saw a required success probability range of approximately 0.30 to 0.39.

Lastly, we had the field positions ranging from 95 to 100. We see here that the highest break-even probability of success is around 0.33 and the lowest is 0.31.

Overall, we see that as the yards-to-go increases within a field position bucket, the break-even probability of success decreases. This is because of the implication that converting on 4th & 10 puts you closer to scoring a touchdown than converting on 4th & 1, which thus results in an increase in your expected points. Therefore, although the risk is higher, as it is empirically less likely to convert in the former scenario than the latter, the payoff is much higher.

It is also interesting to note that the general trend for break-even probability of success increases initially as field position gets closer to scoring, before decreasing again for the last two bins. This makes sense, however, when assessing that the likelihood of making a field goal is fairly low initially. When you're on your opponent's 20/25, it is more likely that you'll make a field goal, and although the odds of subsequently scoring a touchdown after converting are high, it is by no means guaranteed. When you're within 10 yards of scoring, the odds of making a field goal are the highest, but the odds of scoring a touchdown after converting are also quite high.

By using these simulations, coaches can determine when it is worth it to kick a field goal or go for it. It is important to note, however, this analysis should be used with caution as there are many contributing factors that go into a play in real game time. For example, a coach may opt to go for it rather than a field goal if they know their team is stronger than the other team and may get more points than what the simulation provides. Additionally, it can be hard to quantify marginal differences in probability of success, which is why the second table can be used as a supplement. The results between the two tables are generally consistent, showing that teams should go for it when YTG is smaller, generally below 3, or when they are not confident in their ability to make a field goal, such as when they are only on the 40 yard line.