Final Project Report

Levi Johnson

2023-05-05

Problem Statement

The goal of this project is to calculate what are the extremes for an NFL football team to score in a season. More specifically, we want to find what is a once and a season score for an NFL team. Basically what is the theoretical score that we can expect to see only once a season.

Data

The data comes from habitatring.com which had a data set containing all NFL games that have occurred sense the 1999 NFL season to the present. As well the data set initially comes with 45 columns: game_id, season, game_type, week, gameday, weekday, gametime, away_team, away_score, home_team, home_score, location, result, total, overtime, old_game_id, gsis, nfl_detail_id, pfr, pff, espn, away_rest, home_rest, away_moneyline, home_moneyline, spread_line, away_spread_odds, home_spread_odds, total_line, under_odds, over_odds, div_game, roof, surface, temp, wind, away_qb_id, home_qb_id, away_qb_name, home_qb_name, away_coach, home_coach, referee, stadium_id, and stadium.

You can pull this data by running

```
games <- read.csv("http://www.habitatring.com/games.csv")
```

Cleaning Process

To clean the data and for the this project what dropping all the unnecessary columns for this analysis from the dataset. leaving us with just home team, away team, home score, and away score, date. Then each row was split up into two separate rows where each row contains one teams performance in a particular game. So the final data set ends containing: Team, Score, and date.

EDA

Choosing the correct distribution in order to best answer this question is key. The two common distributions used in order to model extremes data is the Generalized extreme value (GEV) and the Generalized Pareo (GP) distributions. They both have slightly different requirements.

In order to fit a GEV model, you fit it on all of the max variable for whatever you are grouping your data based on. In this case the grouping would be based on weeks instead of years or seasons. As there is only 24 years worth of data. So fitting a distribution on only 24 points isn't ideal. So in this case the local maxima are going to be grouped by week, as then we will have 506 vs 24. Which will lead to being able to find more accurate results.

In order to fit a GP model, instead of finding each weeks local maxima a threshold is set which acts as a cut off point which declares whether or not a point is considered an extreme value. The value of this model over

the GEV model is that all extreme events are taken into account. One of the biggest drawbacks of the GEV distribution is that if there are 2 once in a life time games played in the same week. Only one of them will be accounted for however, with a GP model they both will be taken into account. NFL games are unique events and not dependent on each other so it makes more sense fitting a model where that is also the case.

Below is a plot of all scores ever recorded in the NFL sense 1999 plotted vs Time. The X axis is just the score number in order that they where played. The two teams for each game are right next to each other. The Y axis is the score for each team. The blue dots represents if that score is a local maxima. In this case being the highest score by a team during that specific week. The red line is a cut off at the score of 42. While the dashed red line is a cut off at a score of 49.
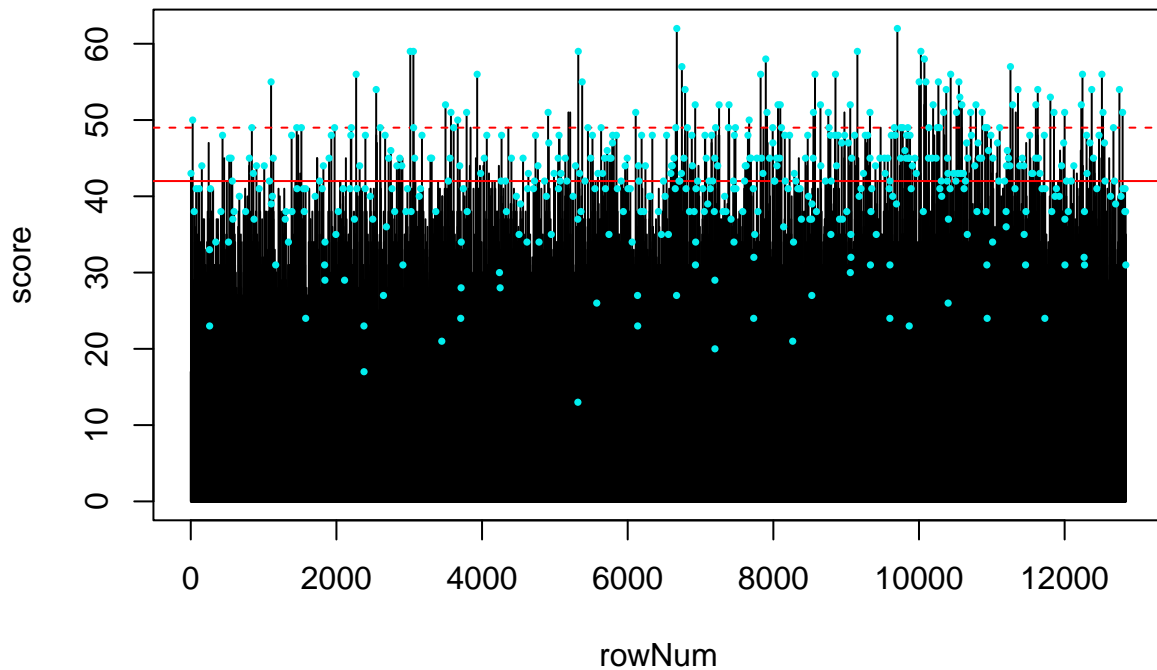
When looking at this plot we can see that the local maxima's are really widely distributed. Which reveals something about our data. Weeks after week 17 or 18 depending on the year(in 2019 the NFL added an additional game to the regular season). Is the playoffs, in which less games are played. Culminating in the Super Bowl where only 1 game is played. In order to properly fit a GEV distribution these playoff games would have to be removed as the sample size for them would be to small. This would require additional data cleaning. That is why we can see some local maxima that are extremely low. There is examples with scores under 20 points being the local maxima which will really skew our GEV model.

Another thing that this model draws attention too is how the scores increase in almost intervals. This can be attributed to the way teams score in football. A touchdown is 6 points and after which most of the time they kick an extra point which will give them an additional point leading to normally scoring 7 points upon a touchdown. As well teams can kick a field goal which will give them an additional 3 points. There is also safeties which occur very rarely but a team can get 2 points from a safety as well. However, points only ever increase in increments with the most common increments being 7 and 3. Resulting in after our cut off at 42 points there is specific lines where you can see lots of scores converge at. The lines being 49 and 45 which is 7 touchdowns for 49 points and six touchdowns and 1 field goal for 45 points. The staggering of the points will most likely skew our final fit to be lower than it should to predict the maxima. However, that is just

the nature of modeling this problem.

Analysis

A GP model was chosen in order to fit to this extremes data. The GP was selected over the GEV due to the small sample sizes to group based on weeks. Additionally potentially significant games that occurred during playoffs would have been forced to be dropped in fitting the GEV. By using the GP distribution we can account for all games that have a significant score.

The threshold chosen for what a significant score should be was 42 points by a team or six touchdowns.

The results of our model fitting:

Histogram of GThresh
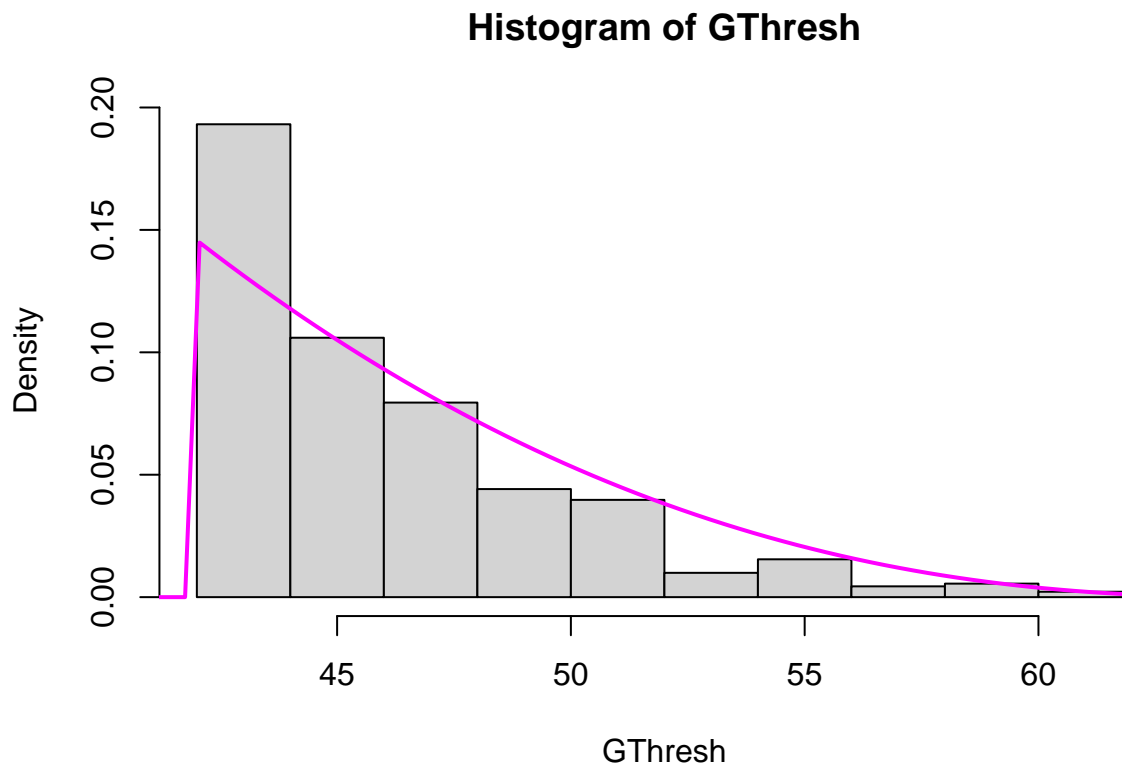
The plot above shows the results of fitting our GP model to the tail of scores data set. The bins in the histogram shows the density of scores occurring at or above 42 points. The pink line is the expected density fit by our GP model. Notably it can be observed that the GP model misses the initial spike at 42 points that we would like it to see. However, by replotting this histogram with 20 bins instead of 10 (see plot below) we can see how there is extremely large spikes at certain numbers notably 42. While there is much less at others such as 43, 46, or 47. This can be attributed to what was discussed above about how football scores work. With certain scores being more common than others.

Histogram of GThresh
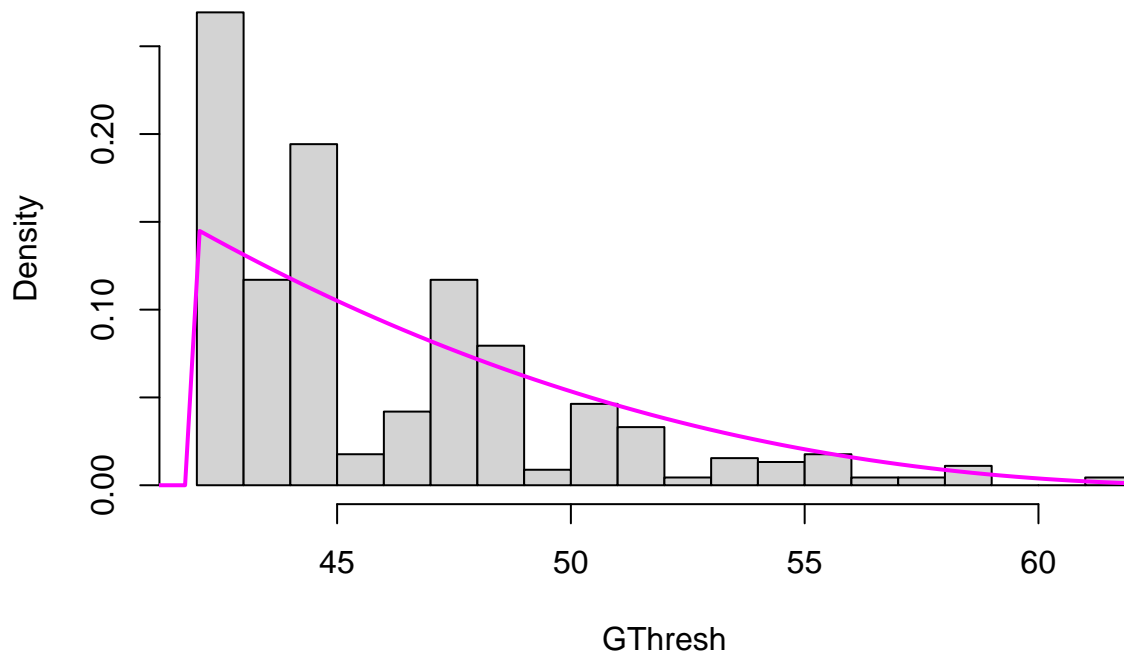
Conclusion

Although this model fails to capture the extreme values at 42. It does capture the tail of the data set really well which is what the goal of this model is. So in order to calculate the once in a season extreme we get a predicted score for once in a season to be 63.042 points. Due to decimal places not being possible in football the rounded value is 63 points. The once in a season score was calculated by finding the 1/535 game. Which 535 is the number of NFL games that occur on average each season.

```
## fevd(x = scores$score, threshold = 42, type = "GP")
##
## [1] "Normal Approx."
##
## [1] "535.083333333333-year return level: 63.042"
##
## [1] "95% Confidence Interval: (53.4268, 72.6562)"
```

Next, calculating a once in 10 season score returned a value of 64 points. Which only returning 1 points more draws some interesting conclusions. First, for a GP if your model returns a negative shape value then your model is accurate only up into a point. So there is a chance that our model is not accurate up to a 1 in a ten season max. As well the confidence intervals for the once in a 10 season max is larger than the 1 in a season max. So it can be reasonably concluded that for the once in the 10 seasons max, that the confidence interval at least is incorrect. However it is worth noting that 63 points is directly divisible by 7 so 63 points is equivalent to 9 touchdowns. Which is a ton for a team to score in a game. However, 64 points would require an additional score so it could potentially be concluded that any score over 9 touchdowns will make the games score a once in 10 seasons event. As a score of 66 points is relatively the same as a score of 64

points as they both are just an additional score over 9 touchdowns. So both of the points listed above are options for this data set.

```
## fevd(x = scores$score, threshold = 42, type = "GP")
##
## [1] "Normal Approx."
##
## [1] "5350.83333333333-year return level: 63.882"
##
## [1] "95% Confidence Interval: (44.9125, 82.8513)"
```

Overall the results of this model are positive and a once in a season max was able to be calculated.

The final Model obtained is:

```
##
## fevd(x = scores$score, threshold = 42, type = "GP")
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  963.02
##
##
##  Estimated parameters:
##      scale      shape
##  6.8628307 -0.3020854
##
##  Standard Error Estimates:
##      scale      shape
## 0.41681562 0.03482626
##
##  Estimated parameter covariance matrix.
##             scale        shape
## scale   0.17373526 -0.012264643
## shape -0.01226464  0.001212869
##
##  AIC = 1930.04
##
##  BIC = 1937.851
```

Future Work

There is lots of potential and directions for future work in this area. Most notably as the NFL continues and more and more games our played the data set will continue to grow and the model will become more accurate. It will be interesting to see what the results might look like in an additional 10 years from now. As well these extremes value models could be carried out to a variety of different stats that occur within football. Notably to the Author Fantasy points could be used in order find what a once in a season fantasy football performance might be. As well rushing yards could be great as the issues caused by incremental score would not apply to that problem. It also would be interesting to apply this same exact model to basketball scores as by increasing by smaller margin could lead to a wider variety of scores and therefore a better fit to the initial tip of the GP model.