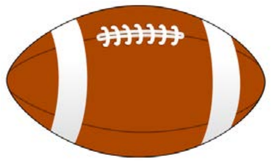
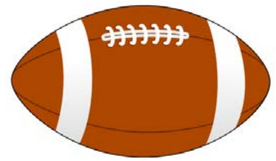


Predicting American Football Games

Alex Crease and Matt Wismer
Computational Bayesian Statistics
Fall 2014



Model 1: Separate Poisson Processes

Our first model considered touchdowns and field goals as separate Poisson processes. By obtaining distributions of both touchdowns per game and field goals per game based on prior data, weighting each distribution and combining them, we obtained a distribution of possible scores per team. Data from each of the teams' records were used to generate team specific Pmfs, and those were used to generate a predictions on games between two teams.

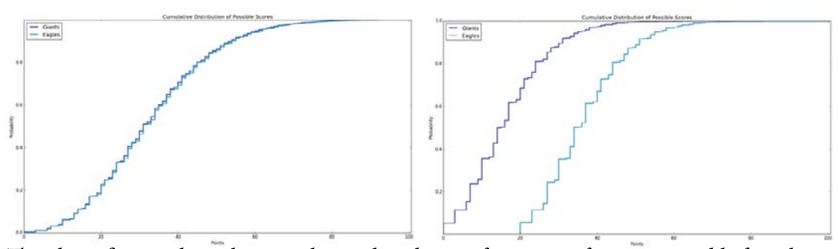
Results

Predictions were made using prior games, and we updated the model at half-time to generate another prediction.

Our model produced the following probabilities:

Time Remaining	Score (E-G)	Eagles Win	Giants Win	Overtime
60	0-0	49.4%	48.2%	2.4%
30	20-0	90.2%	8.2%	1.3%

The following Cdfs were produced based on our distributions:



The above figures show the cumulative distribution functions of points scored before the game started (left) and during half time (right).

Interpretation

The 90% credible interval for the Giants was between 10 and 61. 61 points in a football game is pretty unlikely, and occurs much less than 5% of games. We were a little worried about the broadness of this model, so we rethought our process and created a secondary approach, using a single Poisson Process with the likelihoods of scoring each type of goal mixed in.

Why Game Prediction?

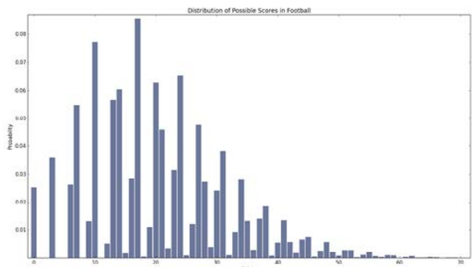
We set out to determine how to predict American football games using Bayesian methods. We chose this problem both because of our enjoyment of American football, and the interesting modelling decisions that we would get to make. Football has a lot of complexities due to different scoring methods and strategy, so we predicted game results using Poisson processes and possible score distributions.

Basic Football Modeling

Our model only incorporates 7 point touchdowns and 3 point field goals for simplicity; all other types of goals are infrequent. We assume that scoring happens in a Poisson manner, so it is equally likely to score at any point in time. We found the probabilities of either team in a match winning or going into OT.

We developed two methods of modeling our system. Both were created using prior data from previous games in the season. We developed specific distributions per team and used them to predict a team's performance as the game progressed.

We used our models to predict a game between the Eagles and the Giants. This would give us a method to compare the two models and validate our results.



A general distribution of points scored during a football match using data averaged across all games this season.

Model 2: Single Poisson Process

In our second model, we considered scoring to be a single Poisson process, with separate distributions of our belief of probability of a team scoring a touchdown versus a field goal. We calculated the probability of scoring a certain number of teach goal based on the binomial distribution, thus goals are related regardless of their type. This model can be thought of as the ability of the offense to get into field goal range, combined with the team's red zone efficiency.

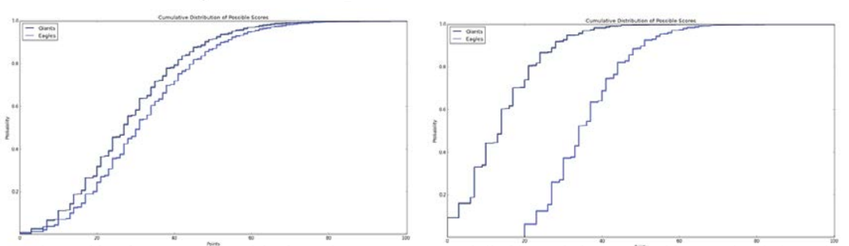
Results

We used the same data as before to compare the two models. We again updated the model at half-time.

Our model produced the following probabilities:

Time Remaining	Score (E-G)	Eagles Win	Giants Win	Overtime
60	0-0	55%	43%	3%
30	20-0	93%	6%	1%

The following Cdfs were produced based on our distributions:

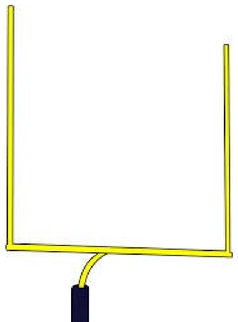


The Cdfs of points scored before the game started (left) and during half time (right) with Model 2. Compared to Model 1, the distributions are tighter yet still similar.

Interpretation

The 90% credible interval for the Giants moved to 7 to 55, so it was a considerably smaller range than previously. However, the credible interval of the Eagles didn't change much; it only went from (10,62) to (10,60). Although this model did tighten our distribution, it is very similar to our first model and thus neither model is significantly better.

Model Expansion



Both models were similar, and it seems that the spread involved was caused by the amount of data that was used to construct the distributions. It would be interesting to see if the model gets more accurate as more games are played in the season. Finding a good way to add more data to the model would be a worthwhile extension.

A main limitation of both of these models is the lack of a defense. As Ray Lewis said in the opening of Madden 2005, "Defense wins championships." By only focusing on the points scored by each team, we ignore information about approximately half the players on the team. Incorporating a defense with each team could be a possible next step with this model.