Applying Glicko Ratings to the NFL

Aaron Weinstock

11/21/2018

This simulation will be based off of the expected game outcome formula allowed by the Glicko rating system. Games remaining in an NFL season will be predicted using this formula.

Relevant History of the Glicko Ratings

Glicko was developed in 1995 by Dr. Mark Glickman, a lecturer at Harvard University, to improve to wellknown Elo rating system (a brief summary at https://en.wikipedia.org/wiki/Elo_rating_system). Though Glicko, like Elo, was originally designed for chess player ratings, it may reasonably applied to team sports, where each "player" is considered a team. In particular, Dr. Glickman's system sought to provide an estimate of a reliability of ratings (essentially, a standard deviation), which is not offered by Elo ratings. Glickman provides the following example in his model documentation to support the need for this improvement:

> *"Suppose two players, both rated 1700, played a tournament game with the first player defeating the second. Under the US Chess Federation's version of the Elo system, the first player would gain 16 rating points and the second player would lose 16 points. But suppose that the first player had just returned to tournament play after many years, while the second player plays every weekend. In this situation, the first player's rating of 1700 is not a very reliable measure of his strength, while the second player's rating of 1700 is much more trustworthy. My intuition tells me that (1) the first player's rating should increase by a large amount (more than 16 points) because his rating of 1700 is not believable in the first place, and that defeating a player with a fairly precise rating of 1700 is reasonable evidence that his strength is probably much higher than 1700, and (2) the second player's rating should decrease by a small amount (less than 16 points) because his rating is already precisely measured to be near 1700, and that he loses to a player whose rating cannot be trusted, so that very little information about his own playing strength has been learned."*

In short, Glicko ratings will assign a high "ratings deviation", or RD, to a player who has not competed in a long time, and a low RD to a player who has competed more recently (i.e., has a lag in their competition). Then, a player's rating as a result of a competition is a function of both his and his opponent's ratings, as well as his and his opponent's RD. For technical details of the rating model and algorithm, visit http://www.glicko.net/glicko/glicko.pdf. A full derivation of the system may be found at http://www.glicko.net/research/glicko.pdf.

The Glicko Expected Outcome Formula

Per the Glicko derivation, expected game outcome is defined as follows:

$$E = \frac{1}{1 + 10^{-g\left(\sqrt{RD_i^2 + RD_j^2}\right) * (r_i - r_j)/400}}$$

Where:

- RD_i and RD_j are the ratings deviations for teams i and j at time t, respectively (for formulas for these quantities, see the above-referenced documentation from Dr. Glickman)
- r_i and r_j are the ratings deviations for teams i and j at time t, respectively (for formulas for these quantities, see the above-referenced documentation from Dr. Glickman)
- $g(u) = \frac{1}{\sqrt{1 + 3q^2(u^2)/\pi^2}}$, where $q = \frac{ln(10)}{400}$

E defines the expected wins (on a 0-1 scale) for player i when competing against player j. In this sense, it can serve as a pseudo win-probability — because it is linearly constrained such that $E_i + E_j = 1$, it defines the likelihood that i beats j.

Why use Glicko for an NFL playoff probability simulation?

There are a number of reasons to apply Glicko to this context; some of the primary reasons are outlined below.

1. Basing ratings (and, more importantly for this context, expected match outcomes) on ratings deviations allows us to account for pauses in game play during NFL bye weeks and NFL playoffs. That this lagged effect will cause for an increase in ratings deviation allows us to attempt a more uncertainty-conscious model using the Glicko ratings.
2. If we define a "rating period" as an NFL week, Glicko ratings for each team can be updated on a week-by-week basis, consequently allowing our expected match outcomes to be dynamic on the same week-by-week basis; this is absolutely requisite for a simulation to live-predict playoff likelihoods.
3. Arguably the most commonly known NFL playoff predictions, done by Nate Silver's FiveThirtyEight site, use an adjusted form of the Elo ratings system (https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/). It could be interesting to compare the results of simulating playoff probabilities based off of these two methods.

What data will be necessary?

Glicko ratings can be calculated in R using the `glicko` function in package `PlayerRatings`. This function requires four data variables for each matchup factoring into the ratings: a ratings period, an identifier for the team i, an identifier for the team j, and a binary result indicating whether or not team i was victorious. This in mind, the following data will be needed.

1. Results from the current NFL season, including week of game, both teams, and the game result (1 for win, 0 for loss, 0.5 for tie).
2. Results from a period prior to this NFL season, from which we can calculate current team ratings (this will be termed the "initialization period"). Again, necessary variables include week, teams, and result, but now must also include season (to distinguish weeks between years).

A combination of week and season will comprise the "rating period" variable (previous years' postseasons will be included in the initialization period as well, with postseason weeks simply considered as additional weeks in a season). "Team identifiers" will simply be team names. Results will be defined as in (1) above.

What parameters must be estimated?

Use of the Glicko rating system requires the estimation of one parameter, c. Per Dr. Glickman, c is a "constant that governs the increase in uncertainty between rating periods." Specifically, c is used to calculate a team's RD at time t by the formula:

$$RD_t = min\Big(\sqrt{RD_{t-1}^2 + c^2},\ 350\Big)$$

Dr. Glickman proposes estimating c by two procedures, both of which are summarized below.

1. "Optimizing predictive accuracy of future games" using the procedure defined in Dr. Glickman's original paper proposing the system (http://www.glicko.net/research/glicko.pdf).
2. "Determine how much time (in units of rating periods) would need to pass before a rating for a typical player becomes as uncertain as that of an unrated player." This would be more of a heuristic, but would require little computational effort.

Refer to `02_Selecting_Default_C.pdf` for details on how a default c was estimated for this particular simulation.