**NHL Game Simulation**

**BACKROUND AND METHODS**

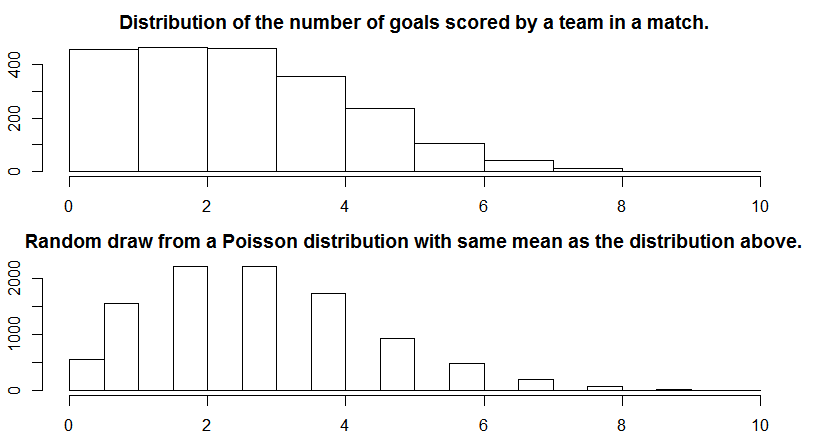
For our simulation, we looked at data from the National Hockey League (NHL) for the current season (2017-18). This included multiple variables for each game played thus far, but for our purposes we cleaned the original data to focus on only the following variables: Date, Home Team, Away Team, Home Goals, Away Goals, and the end result of the game. We then created probabilities using the goals scored by each team in the matches played so far this season. Traditionally, NHL Standings (by points) are scored as followed: two points for a win, one point for an overtime win, and one point for a Shootout loss. For simplicity, we chose to disregard Overtime and Shootout points and simply scored each win as a point. Using these results, we generated stats for each team (by summing Home and Away game stats) to calculate: point totals, goal difference (total goals scored minus total goals conceded) average goals scored, and average goals conceded for the season for each of the 31 teams. Our primary goal was to simulate the outcome of the remaining games left in the 2017-18 season, and to predict the which team is likely to be the top ranked team overall (by points) at end of the regular season. There are multiple ways to predict the result of a hockey game, but for this project we chose to use expected goal values as the determining factor of each game outcome. Here, expected goals are calculated according to the method posted by Mark Taylor. We ran 500 rounds for each of the 31 league teams to produce simulated game results that included total points, goal difference, and ranking per each round.

**SIMULATION AND MODELING**

We chose a Poisson Distribution for our simulation. The Poisson Distribution is a discrete probability distribution that “expresses the probability of a given number of events occurring in a fixed interval of time.” (<https://en.wikipedia.org/wiki/Poisson_distribution>)The Poisson Distribution can be applied when: “(1) the event is something that can be counted in whole numbers; (2) occurrences are independent, (3) the average frequency of occurrence for the time period in question is known; and (4) it is possible to count how many events have occurred”. (<https://www.umass.edu/wsp/resources/poisson/>)

The Poisson Distribution is commonly used to predict outcomes of sport games so we chose to use this method in our project. As noted above, it can be used to count events (in our case events are goals scored) within a specific time period (one hockey game). While there are three periods in a hockey game, for our purposes we considered each single game to be an independent event. We used the number of average expected goal values to determine any given game’s outcome. To get an idea of the Poisson Distribution would be a good fit for our simulation, we modeled the match results for average goals scored and compared that to a model produced by using a Poission Distribution to estimate the likely average amount of goals scored per game (see below).

**Modeling Average Goals Scored per Game: Actual Goals Scored vs. Poisson Distribution Model Goals**



As shown above, a quick comparison between the actual distribution of the average number of goals scored in a game and the modeled Poisson distribution of likely number of average goals scored in a game shows relatively similar results. When comparing the two charts, the actual average number of goals scored relatively closely follows the modeled Poisson distribution. In both graphs, it is less likely to score a higher amount of goals per game than the average goal total, which is around two or three goals per game.

**RESULTS**

Finally, weran our simulation using the expected goal values we determined as the lambda for the distribution. While this simulation has only modeled regular season games, we can now see which team has the highest probability of be ranked first going into the playoffs. From our simulation, it looks like the Tampa Bay Lightning are the current favorite at a rate of 83.4%. At this point in the actual season, Tampa Bay has the best overall record of 48 wins and 18 losses. Using our model, we can also look at all each individual team's probabilities to a respective ranking out of 31 total teams.

**IMPROVEMENTS TO MODEL**

There are a few improvements we felt could be made to our model going forward. It would have been ideal to score the team rankings (determined by points) more accurately by including Overtime and Shootout data. It appears that the probabilities are heavily influenced by the current overall win and loss records at this point in the season.

* Games are not really “Independent events” in reality….
  + Home court advantage
  + Past few game steak/records
  + Rivalries, etc
  + Travel distances?
* hockey is traditionally a low scoring game