Decision Making in American Football: Evidence from 7 Years of NFL Data

Konstantinos Pelechrinis

School of Information Sciences
University of Pittsburgh
kpele@pitt.edu

Do team managers and coaches make rational decisions? In the era of big data one would expect coaching decisions to always be rational with respect to the maximization of the expected points scored. In order to examine this hypothesis we use data from the National Football League for the past 7 seasons to analyze two specific discrete decisions/choices that coaches face several times during a game. In particular, we analyze (i) the Point(s) After Touchdown (PAT) and (ii) the fourth down decisions. We use mean-field approximations and find that despite the statistical evidence available to them through years worth of data, coaches tend to make the status quo decisions, which in the majority of the cases does not lead to point maximization. One of the possible explanations that we put forward for this behavior is the fact that their coaching objective might be to minimize the variance of the expected points scored.

1 Introduction

While American football is viewed mainly as a physical sport - and it surely is - at the same time is probably the most strategic sports game, a fact that makes it appealing even to international crowd [14]. This has led people to analyze the game with the use of data analytics methods and game theory. For instance, after the controversial last play call of Super Bowl XLIX the Economist argued with data and game theory that this play was rational and not that bad after all [10]. Despite the fact that statistical analysis has always been part of the sports industry, the recent ability to analyze and collect large volumes of data (both traditional boxscore as well as sensor data) has led to the emergence of an evidence-based approach in various sports.

For example, Clark *et al.* [7] analyzed the factors that affect the success of a field goal kick and contrary to popular belief they did not identify any situational factor (e.g., regular vs post season, home vs away etc.) as being significant. In another direction the authors in [15] and [20] studied models and systems for determining the points scored in NFL games. The impact of the much-discussed off-field misconduct of NFL players was studied by Stair *et al.* [19] who showed that it does not affect a team's performance. In a different sport, Fewell *et al.* [11] analyzed data from the 2010 NBA play-offs using network theory. In particular, they considered a network where each team player is a node and there is an edge between two players if they exchanged a pass. Using this structure they found that there is a consistent association between a team's advancement in the next playoff round and the values of clustering and network entropy. Bar-Eli *et al.* [6, 5] further studied decision making in soccer penalty kicks. In particular, they collected information from 286 penalty kicks from professional leagues in Europe and South America and analyzed the decisions made by the penalty tackers and the goal keepers. Their main conclusion is that from a statistical standpoint, it seems to be more advantageous for a goal keeper to defend by remaining in the goal's center. Furthermore, Di *et al.* [9] analyzed the motion of 200 soccer players from 20 games of the Spanish Premier League and 10 games of the Champions League and found that the different positional roles demand for different work intensities. In another direction the authors in [8] analyzed the pass behavior of rugby players. They found that the time period required in order to close the gap between the first attacker and the defender explained 64% of the variance found in pass duration and this can further yield information about future pass possibilities. Similar analyses has been performed for the pass behavior of NHL teams [3]. Data from professional sporting competitions have also been used extensively as a proxy for testing various hypotheses of firm optimization as well as for understanding the way people behave, understand and misinterpret statistical information (e.g., [17, 12, 4]). This implies that there is much to be learned by analyzing coaching decisions.

Despite these academic studies the question on whether coaches learn from historic information remains open. In particular, the goal of every team is to win, which - in a good approximation - is achieved by maximizing the number of points scored. Hence, one should expect the decisions that the coaches are making to always be rational with respect to this objective. However, it seems that rational decisions made by the coaching staff are many times received with questions and are scrutinized as the article from the Economist reveals. Such scrutiny might discourage coaches from following rational decisions especially when their justification is complex. Other times, as is the case with a Texas high school football coach that never punts [1], their strategy is obtained with surprise despite the success. The *status quo bias* [13] can further perpetuate this behavior and hence, coaches make decisions that differ from those that a rational agent would take. Using game data from the National Football League for the period between 2009 and 2015 we focus on two particular decisions that the coaches face multiple times during the course of a single game, namely, the Point(s) After Touchdown (PAT) and the fourth down decisions. The status quo for PAT is to go for an extra point kick unless if specific situations appear, while the status quo for fourth down decisions is to punt, unless again if specific situations appear (e.g., game clock running out and trailing in the score). Our game data analysis shows that actually both of these decisions, in the majority of the cases, are not rational when the objective is to maximize the expected number of points scored in a game. This tenacity of NFL coaches is rather surprising especially given that similar issues have been reported in the literature since 1967 [16, 18]! One might have expected that with data analysis and statistics being an integral part of sports organizations this behavior would have changed. Finally, trying to shed light on this coaching behavior, we argue that the decisions made by the coaches is in accordance with a strategy of minimizing the variance of the expected points scores.

2 Analysis

NFL Dataset: In order to perform our analysis we utilize a dataset collected from NFL's Game Center for all the regular season games between the seasons 2009 and 2015. We access the data using the Python `nflgame` API [2]. The dataset includes detailed play-by-play information for every game that took place during these seasons. In total, we collected information for 1,792 regular season games. In what follows we analyze the decisions made with respect to the PAT and fourth down.

2.1 Points After Touchdown

Once a team scores a touchdown (worth 6 points) it has the option to either kick an extra point field goal from the 15-yard line or make a regular play from the 2-yard line for 2 points, namely, a two-point conversion. In some cases the decision is easy (i.e., the team is trailing by 2 points with the clock running out) but in most of the touchdowns the decision is not so clear since special circumstances do not exist. In fact, one would expect that such extraordinary situations appear only for the touchdowns towards the end of the game. Analyzing our data we find that the dominant decision is to settle for the extra point after a touchdown instead of trying to score more points and attempting a two-point conversion. In particular, from the 9,021 touchdowns in our dataset, only 460 of them were followed by a two-point conversion attempt. From these, 235 were successful, that is, an overall 51% success rate. On the other hand, from the 8,561 extra point kick attempts, 8,425 were successful, which translates to a 98.4% success rate. With $s_{\texttt{2pts}}$ and $s_{\texttt{kick}}$ being the success rates for the two-point conversions and extra point kicks respectively, the expected point differential benefit per touchdown $\mathbf{E}[p]$ of a two-point conversion over an extra point kick is given by:

$$\mathbf{E}[p] = 2 \cdot s_{\texttt{2pts}} - 1 \cdot s_{\texttt{kick}} \tag{1}$$

With the extra point kick, the probability of success is 0.984 and thus, the expected number of points is 0.984. On the contrary, with the two-point conversion the success rate is only 0.51 but the expected number of points is 1.02, which ultimately gives a positive expected point differential benefit (i.e., $\mathbf{E}[p] > 0$). Of course, the net gain per touchdown is fairly small and not all the teams have the same success rate in the two-point conversions and the extra point kick attempts. Hence, not all teams would necessarily benefit from this strategy. Figure 1 presents the expected benefit $\mathbf{E}[p]$ per team using the corresponding success rates from our 7-year dataset.

One interesting aspect of this analysis is that the PAT rules changed at the beginning of the 2015 season. In particular, the extra kick is snapped from the

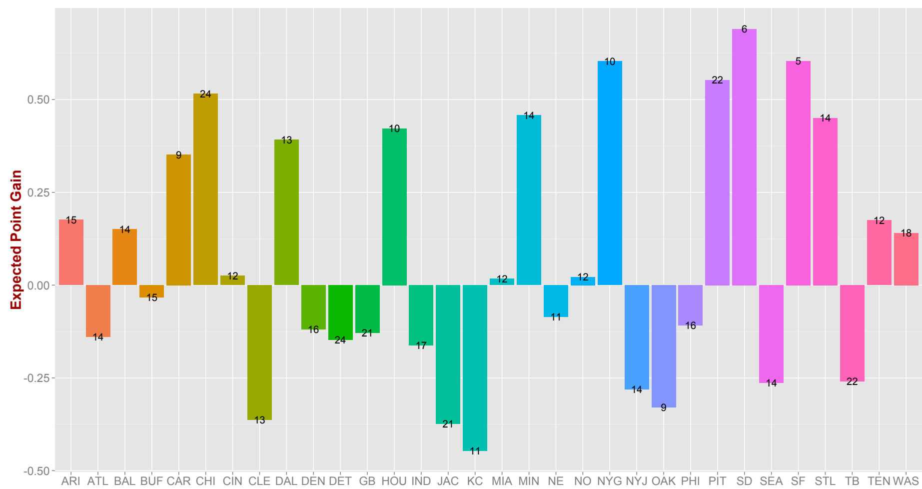

Fig. 1: Different teams have different expected point gains based on the corresponding success rates $s_{\texttt{2pts}}$ and $s_{\texttt{kick}}$. On the top of each bar we also present the number of times that the corresponding team attempted a two-point conversion in our dataset.

15-yard line instead of the 2-yard line that used to be all the previous years. This change has led to a statistically significant reduction in $s_{\texttt{kick}}$. Table 1 depicts the success rates for the extra point kicks over the years that our dataset spans. As we can observe there is a significant drop of approximately 5% (p-value $< 10^{-6}$) in the success rate during the 2015 season! On the contrary, there was no impact on the success rate of the two-point conversion attempts. Overall, statistically comparing the median yearly expected points with the two strategies, the two-point conversion provides a clear benefit (p-value < 0.1).

Year	2009	2010	2011	2012	2013	2014	2015
$s_{\texttt{kick}}$	0.9814	0.9884	0.9942	0.9935	0.9960	0.9926	**0.9416**
$s_{\texttt{2pt}}$	0.4426	0.5471	0.5	0.55	0.4929	0.5161	0.5247

Table 1: The new PAT rules introduced in 2015 have led to significant drop of the success rate of the extra point kick by approximately 5% (p-value < 0.001).

Currently, the decision to attempt a two-point conversion depends mainly on the score differential and the time remaining in the game. For instance, when a team scores a touchdown towards the end of the game giving them the lead by one point, it is typical to attempt a two point conversion that will potentially give a 3 point lead, hence, putting the pressure on the opposing team to score a touchdown to win the game. What our analysis suggests is that these factors should not impact the decision to attempt a two point conversion, since the expected point benefit is larger regardless of these factors. Even if we consider that the nominal benefit is small, one would have expected teams to attempt for a two-point conversion way more often than the 5% they currently do.

2.2 Fourth Down Decisions

Another decision that coaches have to take - more often than the PAT - is related with the fourth down situations. The teams have 4 tries to advance 10 yards on the field. If they fail to do so the opponent takes the ball at the yard line that the team was stopped. The teams have to make a choice after the first 3 tries on whether to go for their fourth try and keep their drive alive or whether to punt the ball and push the ensuing drive of the opposing team further from their own goal line. Depending on the distance to the goal they might also have the possibility to try for a field goal for 3 points.

In the vast majority of the cases that coaches face this decision, they decide to either punt or kick a field goal. Exceptions of course appear in specific situations, e.g., when a team is trailing by more than 3 points and the clock is running down. Using mean field approximations we calculate the net benefit from "going for it" on fourth down. In order to estimate the expected benefit we need to compute from our data the following quantities: (i) the conversion rate of a fourth down conversion (as a function of both the field position and the yards to cover for a first down), (ii) the success rate of a field goal (as a function of the distance from the goal) and (iii) the probability of success for a drive (field goal or touchdown) as a function of the starting field position of the drive. The latter is needed in order to calculate the (average) impact that a failed fourth down conversion will have on the ensuing drive of the opposing team.

Fourth down conversion rate: We begin by examining the success rate of the fourth down conversion attempts. Overall, the fourth down conversion rate is a stunning 77.9%! Furthermore, we examine whether this conversion rate is affected by factors such as the position of the offense on the field and the yards remaining for a first down. We first examine the impact of the field position on the conversion rate. As one might have expected (figure omitted due to space limitation) the success rate of the fourth down conversion is not impacted by the field position.

However, one might expect that the fourth down conversion rate is affected by the yards needed to be covered. In particular, the shorter yardage intuitively has a higher success rate as compared to longer yardage. Figure 2 depicts our results, which verify this intuition. In fact, there is a declining trend; the more yards the offense has to cover in the fourth down, the lower the chances of a successful conversion. An interesting observation is the fact that 55% of the fourth down attempts in our dataset are in fourth down and one situations, i.e., the offense has to cover only one yard. Hence, the overall fourth down conversion rate is skewed, since the maximum conversion rate is observed in these situations and is equal to 89%. Furthermore, the vast majority of the attempts (95% of

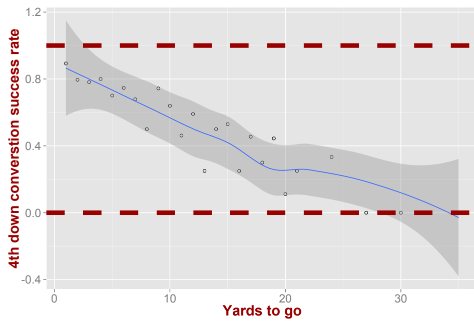

Fig. 2: The fourth down conversion rate reduces with an increase of the yardage left for the first down.

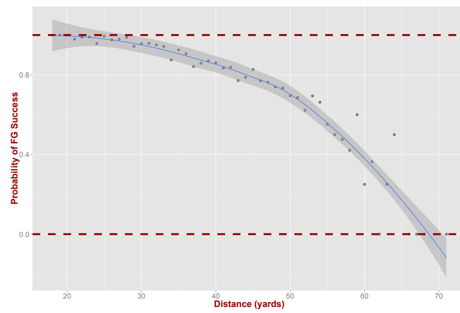

Fig. 3: The overall success rate of a field goal attempt is 85% and declines with the distance from the goal.

them) require at most 10 yards to reach the first down mark. Adjusting for this, we obtain an estimate for the average fourth down conversion rate of 73%.

Field Goal Success Rate: When teams face a fourth down decision in their opponent's territory they can decide to settle for a field goal, which will provide them with 3 points if successful. In order to calculate the expected payoff from a field goal attempt we calculate the success rate of the kick as a function of the distance from the goal. Figure 3 depicts our results. As we can see there is a slowly declining field goal success rate, which exhibits a steeper decline after the 50 yards. However, only 11% of the field goal attempts in our dataset had a distance larger than 50 yards. These typically correspond to efforts to tie or win a game when the clock is running down; under regular circumstances teams would most probably have punted the ball. Overall, the success rate of a field goal (not controlling for the distance from the goal) is 85.5%.

Success of a Drive and Starting Field Position: Turning the ball on downs does not only impact the current drive of the offense by terminating it with 0 points, but it might also give the opponent a very good starting field position. After a touchdown or a field goal the team that scored, kicks the ball from their own 35 yard line. The opponent has a dedicated player (called kick returner) who receives the ball and tries to advance it on the field. Typically, most of the kicking teams attempt to kick the ball beyond the opponent's goal line, which will not give the opportunity to the kick returner to advance the ball. In this case the next drive starts from the offense's own 20 yard line.

The starting position of a team can potentially impact the success of the drive. Hence, in order to calculate the potential loss from a failed fourth down conversion we need to estimate how a turn on downs will impact the success of the ensuing drive from the opposing team. Figure 4 presents the fraction of drives that resulted in a field goal, touchdown or failed (i.e., ended with turnover or punt) as a function of the starting position captured by the distance to be covered for a touchdown. As we can see when the distance that the offense has to cover at the beginning of the drive is less than 25 yards the probability of scoring a touchdown is rapidly increasing, while the probability of not scoring at all reduces rapidly as well. However, both of them are much smaller compared to the probability of a failed drive (for a distance greater than 50 yards).

Clearly the starting field position can impact the success of the drive especially when this drive starts at the opponents territory. This might be the reason that an offense rarely goes for fourth down when they are in their own territory; a failed conversion increases dramatically the chances for a scoring ensuing drive from the opponent. However, when the offense has entered the opponents territory (i.e., distance to the goal is less than 50) failing to convert on a fourth down increases the chances of a scoring ensuing drive by the opponent by only 7% on average as compared to the baseline case of a touchback.

Mean Field Net Point Benefit Using the above results we will estimate the mean field approximation of the net gain for attempting the fourth down conversion. For the mean field approximation of the net point gain $\mathbf{E}[P]$ we need to calculate the expected point benefit $\mathbf{E}[P^+]$ from a successful fourth down conversion as well as the expected point cost $\mathbf{E}[P^-]$ from a potential failed conversion. The expected benefit $\mathbf{E}[P^+]$ is a function of the fourth down conversion rate $s_{\mathtt{4conv}}$ as well as the field position l, i.e., $\mathbf{E}[P^+] = f_1(s_{\mathtt{4conv}}, l)$. On the contrary, the expected point cost $\mathbf{E}[P^-]$ is a function of the success rate of a field goal $s_{\mathtt{fg}}$, which itself is a function of the field position l, and the increase in the probability $\Delta\pi_{\mathtt{td}}$ and $\Delta\pi_{\mathtt{fg}}$ of the ensuing opponent's drive leading to a touchdown or field goal score respectively, which itself depends on the field position l as well, i.e., $\mathbf{E}[P^-] = f_2(s_{\mathtt{fg}}, \Delta\pi_s, l)$. In particular,

$$\mathbf{E}[P^+] = 6 \cdot s_{\mathtt{4conv}}^{\gamma(l)} \tag{2}$$

$$\mathbf{E}[P^-] = 3 \cdot s_{\mathtt{fg}} + (3 \cdot \Delta\pi_{\mathtt{fg}} + 6 \cdot \Delta\pi_{\mathtt{td}}) \tag{3}$$

where $\gamma(l)$ is the number of times that the offense will need to convert a fourth down to reach the goal line. Clearly the further from the goal line (i.e., small l) the larger the expected value of $\gamma(l)$. In order to have a realistic estimate for $\gamma(l)$ we analyzed all the approximately 43,000 drives from all the games in our dataset. The average drive length is 29 yards. This means that, on average in order to keep the drive alive, the team will need to convert on fourth down once every 29 yards. Therefore, if a team starts at its own 20-yard, this means that they will have to successfully convert on average 2.7 times before reaching the goal line. In general, with l being the starting field position of a team (i.e., the yards to cover are $100 - l$) we have $\gamma(l) = \frac{100-l}{29}$. One of the parameters that is hard to estimate is the $s_{\mathtt{4conv}}$. As alluded to above the observed fourth down attempts are skewed towards scenarios that are more probable to success (i.e., small yardage to go). In order to obtain a good estimate for our mean field estimation, and given that we are interested in the average case, we used the drives in our dataset and calculated the yards-to-go at the end of every drive. The mean value for the yardage is 7.58, which translates from Figure 2 to $s_{\mathtt{4conv}} = 0.67$.

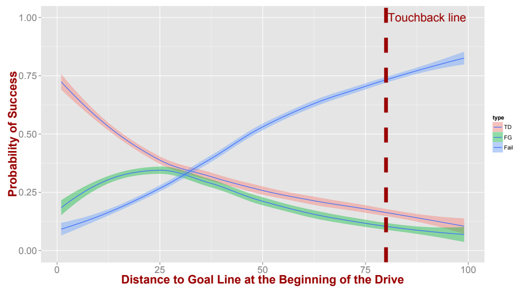

Fig. 4: A drive that starts in the opponents territory has significantly increased probability of leading to a scoring play (touchdown or field goal).

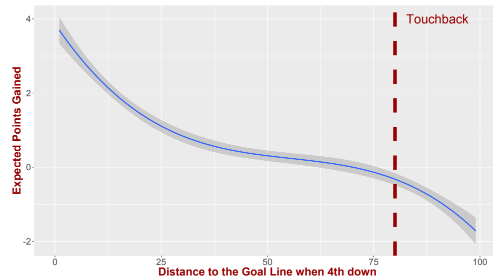

Fig. 5: The expected point benefit is positive for most of the field. The benefits become particularly tempting in the opponent's territory.

Using the results from Figures 2, 3 and 4 we obtain Figure 5. This figure depicts the expected point benefit (i.e., $\mathbf{E}[P] = \mathbf{E}[P^+] - \mathbf{E}[P^-]$) that the offense will have as a function of the field position at the first time it faces a fourth down situation. In particular, the horizontal axis represents the distance to the goal line. As we can see, $\mathbf{E}[P]$ is positive for more than 65% of the field. In fact the average point gain is 0.54 points per drive (p-value < 0.0001), which can translate to significant point gains over the course of a game. In conclusion, even though there clearly are cases where going for it on fourth down does not provide any benefit - i.e., in the regime where $\mathbf{E}[P] < 0$ - teams are extremely reluctant on attempting the conversion even though the game data show that there can be significant point benefits.

3 Discussion, Limitations and Conclusions

In this study we have used data from the last 7 NFL seasons to examine the decisions made by coaches in two specific play calls that teams face very often in each game, namely, PAT and fourth down. We would like here to emphasize on the fact that our analysis simply provides evidence that from the perspective of point maximization the coaches do not act rationally. More specifically, we do **not** claim that it is always beneficial to attempt a two-point conversion. There are cases where an extra point kick is clearly the optimal choice. For instance, if a team is down by 6 points and scores a touchdown to tie the game with 1 second left on the clock, it is obvious that the optimal strategy is to go for an extra point kick since the probability of success (approximately 95%) is much higher as compared to converting the two-point attempt (a little over 50%). However, for the vast majority of the game (excluding possibly the final part of the game) the decisions for maximizing the expected points should also be maximizing the probability of winning.

One of the reasons behind the current decisions made from the coaching staff is potentially their risk-averse attitude. In particular, while providing higher expected point benefit, both the two-point conversion and the decision to attempt a fourth down, exhibit higher variance as well. Therefore, it can be the case that the objective of a team's coaching staff is to minimize the variance for the expected points scored, rather than maximize the latter. In particular, the standard error for the two-point conversion is 2.3%, while the standard error for the extra point kick is only 0.01%! Similar is the situation with the fourth down conversion rate (standard error is 4.7%) and the field goal success rate (standard error is 0.42%).

Therefore, we believe that teams will continue to be conservative, mainly due to the fact that in order for the expected outcome to converge to what is predicted a very large number of attempts need to be made. For instance, with respect to the two-point conversion in a single game a very small number of two-point conversion attempts will be made by each team (on average 2.5 touchdowns per team per game). In fact even teams that have started adapting their game plan - mainly due to the PAT rule changes - are still overly conservative. For instance, the Pittsburgh Steelers during the 2015 season attempted multiple two-point conversion in "unconventional" times of the game (e.g., during the first quarter, etc.). However, they only attempted 11 conversion (converting 8 of them) out of the 45 touchdowns they scored, which is an overall rate of just about 25%.

One of the limitations of our current analysis is the fact that we are mainly focused with decisions regarding the offensive unit of a team. However, the defensive unit can have a significant impact on the chances of a team winning the game. Therefore, the objective of the coaching staff could potentially include maximizing the point differential between the two teams. In this case, the decision on going for 4th down also depends on the confidence level for the defensive unit stopping the ensuing drive. Our results (Figure 4) present the average case - aggregate over all teams and seasons. A team with a low-ranked defensive unit might prefer to pin the ball deep into the opponents territory through a punt in an effort to keep the point differential constant (i.e., through stopping the upcoming opponent's drive).

Finally, as implied above, our analysis for the fourth down decision has an inherent bias at the computation of $s_{\texttt{4conv}}$. The observed attempts for conversion are not a random sample of all fourth down situations but rather correspond to the cases that the coaches believe the conversion is doable. Therefore, the actual $s_{\texttt{4conv}}$ is highly probable to be lower. However, we would like to emphasize here that in our results we have not used the average fourth down conversion rate but rather the conversion rate for the average yardage to go (that is 7.58) at the end of a drive. Of course, even more detailed decision guidance could be obtained if one computes $\mathbf{E}[P]$ as a function of the yardage to cover y with the fourth down instead of simply using the average yardage to go at the end of every drive. In the future we plan on identifying better ways to obtain an unbiased estimate for the $s_{\texttt{4conv}}$.

References

1. The high school football coach that never punts. http://fivethirtyeight.com/features/the-high-school-football-coach-who-never-punts/ (2012), accessed: 2016-01-08
2. Nfl game center api. https://github.com/BurntSushi/nflgame (2012), accessed: 2016-01-12
3. Nhl passing project. https://hockey-graphs.com/2015/12/18/passing-project-data-release-volume-i/ (2015), accessed: 2016-01-08
4. Albright, S.C.: A statistical analysis of hitting streaks in baseball. Journal of the American Statistical Association 88(424), 1175–1183 (1993)
5. Bar-Eli, M., Azar, O.H.: Penalty kicks in soccer: an empirical analysis of shooting strategies and goalkeepers' preferences. Soccer & Society 10(2), 183–191 (2009)
6. Bar-Eli, M., Azar, O.H., Ritov, I., Keidar-Levin, Y., Schein, G.: Action bias among elite soccer goalkeepers: The case of penalty kicks. Journal of Economic Psychology 28(5), 606–621 (2007)
7. Clark, T., Johnson, A., Stimpson, A.: Going for three: Predicting the likelihood of field goal success with logistic regression. In: The 7th Annual MIT Sloan Sports Analytics Conference (2013)
8. Correia, V., Araujo, D., Craig, C., Passos, P.: Prospective information for pass decisional behavior in rugby union. Human Movement Science 30(5), 984 – 997 (2011)
9. Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F., Bachl, N., Pigozzi, F.: Performance characteristics according to playing position in elite soccer. International journal of sports medicine 28(3), 222 (2007)
10. Economist, T.: Game theory in american football: Defending the indefensible. http://www.economist.com/blogs/gametheory/2015/02/game-theory-american-football (2015), accessed: 2016-01-12
11. Fewell, J., Armbruster, D., Ingraham, J., Petersen, A., Waters, J.: Basketball teams as strategic networks. PLOS ONE 10.1371/journal.pone.0047445 (2012)
12. Gilovich, T., Vallone, R., Tversky, A.: The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology 17(3), 295 – 314 (1985)
13. Kahneman, D., Knetsch, J.L., Thaler, R.H.: Anomalies: The endowment effect, loss aversion, and status quo bias. Journal of Economic Perspectives 5(1), 193–206 (1991)
14. Lamb, C., Hair, J., McDaniel, C.: Essentials of Marketing. ISBN-13: 978-0538478342 (2012)
15. Pfitzner, C., Lang, S., Rishel, T.: Factors affecting scoring in nfl games and beating the over/under line. The Sport Journal (2014)
16. Porter, R.C.: Extra-point strategy in football. The American Statistician 21(5), 14–15 (1967)
17. Risen, J., Gilovich, T., Thaler, R.: Sudden death aversion: Imagining the worse and avoiding it. The University of Chicago Booth School of Business (Technical Report) (2006)
18. Sackrowitz, H.: Refining the point (s)-after-touchdown decision. Chance 13(3), 29–34 (2000)
19. Stair, A., Day, A., Mizak, D., Neral, J.: The factors affecting team performance in the nfl: does off-field conduct matter? Economics Bulletin 26(2), 1–9 (2008)
20. Warner, J.: Predicting margin of victory in nfl games: Machine learning vs. the las vegas line. Technical Report (2010)