American Football

Hal S. Stern

Department of Statistics, Iowa State University

December 24, 1997

1 Introduction

1.1 A brief description of American football

There are a number of sports played across the world under the name *football*, with most of the world reserving that name for association football or soccer. Soccer is discussed in Chapter 5 of this volume, this chapter considers American football, the basic structure of which is described in the next paragraph. Most of the material in this chapter is discussed in terms of the National Football League, the professional football league in the United States. Other versions of the game include college football in the United States and professional football in Canada. There can be substantial differences between different versions of American football, e.g., the Canadian and United States games differ with respect to the length of the field and the number of players among other things. Despite these differences, the methodology and ideas discussed in this chapter should apply equally well to all versions of American football.

American football, which we shall call football from this point on, is played by two teams on a field 100 yards long with each team defending one of the two ends of the field (called goal lines). Games are 60 minutes long, and are broken into four 15-minute quarters. The two teams alternate possession of the ball and score points by advancing the ball (by running or throwing/catching) to the other team's goal line (a touchdown worth 6 points with the additional opportunity to attempt a one-point or two-point conversion play), or

failing that by kicking the ball through goal posts situated at the opposing team's goal (a field goal worth 3 points). The team in possession of the ball (the offense) must gain 10 or more yards in four plays (known as downs) or turn the ball over to their opponent. The ball is advanced by running with it, or by throwing the ball to a teammate who may then run with the ball. As soon as 10 or more yards are gained, the team starts again with first down and a new opportunity to gain 10 or more yards. If a team has failed to gain the needed 10 yards in three plays then it has the option of trying to gain the remaining yards on the fourth play or kicking (punting) the ball to its opponent to increase the distance that the opponent must move to score points. This very fast description ignores some important aspects of the game (the defense can score points via a safety by tackling the offensive team behind its own goal line, teams turn the ball over to their opponents via dropped/fumbled balls or interceptions of thrown balls) but should be sufficient for reading most of this chapter.

1.2 A brief history of statistics in football

As with the other sports in this volume, large amounts of quantitative information are recorded for each football game. These are primarily summaries of team and individual performance. For United States professional football these data can be found as early as the 1930s (Carroll, Palmer, and Thorn, 1988) and in United States collegiate football they date back even earlier. We focus on the use of probability and statistics for analyzing and interpreting these data in order to better understand the game, and ultimately perhaps to provide advice for teams about how to make better decisions.

The earliest significant contribution of statistical reasoning to football is the development of computerized systems for studying opponents' performances (e.g., Purdy (1971) and Ryan et al. (1973)). Professional and college teams currently prepare reports detailing the types of plays and formations favored by opponents in a variety of situations. The level of detail in the reports can be quite remarkable, e.g., they might indicate that Team A runs to the

left side of the field 70% of the time on second down with five or fewer yards required for a new first down. These reports clearly influence team preparation and game-time decision-making. These data could also be used to address strategy issues (e.g., should a team try to maintain possession when facing fourth down or kick the ball over to its opponent) but that would require more formal analysis than is currently done – we consider some approaches later in this chapter. It is interesting that much of the early work apply statistical methods to football involved people affiliated with professional or collegiate football (i.e., players and coaches) rather than statisticians. The author of one early computerized play-tracking system was 1960s professional quarterback Frank Ryan. Later in this chapter we will see contributions from another former quarterback, Virgil Carter, and a college coach, Homer Smith.

1.3   Why so little academic research?

Football has a large following in the United States and Canada, yet the amount of statistical or scientific work by academic researchers lags far behind that done for other sports (most notably, baseball). It is interesting to speculate on some possible causes for this lack of results. We briefly describe three possible causes: data availability, the nature of the game of football, and professional gambling.

First, despite enormous amounts of publicity related to professional football, it is relatively difficult to obtain detailed (play-by-play) information in computer-usable form. This is not to say that the data don't exist – they clearly do exist and are used by teams during the season to prepare their summaries of opponents' tendencies. The data have not been easily accessible to those outside the sport. The quality of available data is improving, however, as play-by-play listings can now be found on the World Wide Web through the National Football League's own site (http://www.nfl.com). These data are not yet in convenient form for research use.

3

A second contributing factor to the shortage of research results concerns the nature of the game. Examples of the kinds of things that can complicate statistical analyses: scores occur in steps of size 2,3,6,7,8 rather than just a single scoring increment, the game is time-limited with time management an important part of strategy, actions (plays) move the ball over a continuous surface with an emphasis on 10-yard pieces. All of these conspire to make the number of possible situations that can occur on the football field extremely large which considerably complicates analysis.

One final factor that appears to have worked against academic research about the game itself is the existence of a large betting market on professional football games. A great deal of research has been carried out on the football betting market including methods for rating teams (described later in the chapter) and methods for making successful bets (described elsewhere in the book). Unfortunately, because of its applicability to gambling, a large portion of this research is proprietary and unavailable for review by other researchers.

The remainder of this chapter is organized as follows. The next three sections describe research results and open problems in three major areas: player evaluation, models for assessing football strategy, and rating teams. Following that is a section that touches on other possible areas of research. The chapter concludes with a brief summary and a list of references. One reference merits a special mention at this point; *The Hidden Game of Football*, a 1988 book by Bob Carroll, Pete Palmer, and John Thorn, is a sophisticated analysis of the game by three serious researchers with access to play-by-play data. Written for a popular audience, the book does not provide some of the details that readers of this book (and the author of this chapter) would find interesting. We refer to this book quite often and use CPT (the initials of the three authors) to refer to it.

2 Player evaluation

Evaluation of football players has always been important for selecting teams and rewarding players. Formally evaluating players, however, is a difficult task because several players contribute to each play. A quarterback may throw the ball five yards down the field and the receiver, after catching the ball, may elude several defensive players and run 90 additional yards for a touchdown. Should the quarterback get credit for the 95-yard touchdown pass or just the 5-yards the ball traveled in the air? What credit should the receiver get? We first review the current situation and then discuss the potential for evaluating players at several specific positions.

2.1 The current situation

Evaluation of players in football tends to be done using fairly naive methods. Football receivers are ranked according to the number of balls they catch. Running backs are generally ranked by the number of yards they gain. Punters are ranked according to the average distance they kick the ball without regard to whether they are effective in making the opponent start from poor field position. Kickers are often ranked by the number of points scored. The most complex system, the system for ranking quarterbacks, is quite controversial – we review it shortly. Defensive players receive little evaluation in game summaries beyond simple tallys of passes intercepted, fumbles recovered, or quarterbacks tackled for a loss of yardage. Offensive linemen, whose main job is to block defensive players, receive essentially no formal statistical evaluation. Several problems are evident with the current situation: the best players may be misevaluated (or not evaluated at all) by existing measures, and it can be very difficult to compare players of different eras (because there are different numbers of games per season, different football philosophies, and continual changes in the rules of the game).

2.2 Evaluating kickers

The difficulty in apportioning credit to the several players that contribute to each play has meant that a large amount of research has focused on aspects of the game that are easiest to isolate, such as kicking. Kickers contribute to their team's scoring by kicking field goals (worth three points) and points-after-touchdowns (worth one point). On fourth down a coach often has the choice of: (1) attempting an offensive play to gain the yards needed for a new first down; (2) punting the ball to the opposition, or (3) attempting a field goal. Evaluation of the kicker's ability will have a great deal of influence on such decisions. Berry and Berry (1985) use a data-analytic approach to estimate the probability that a field goal attempted from a given distance will be successful for a given kicker. They then propose a number of measures for comparing two kickers, e.g., the estimated probability of converting a 40-yard field goal. Interestingly, Morrison and Kalwani (1993) examine all professional kickers and conclude that binomial variability is sufficiently large that a null model that says *all* kickers are equally good (or bad) would be accepted. This suggests that rating kickers may not be a good idea at all. Of course, as they point out it is also possible that some kickers are indeed better than others but that the 30 or so field goal attempts per season are not enough to detect the difference. In addition to comparing kickers, it can be valuable to explore the factors affecting the probability of success of a field goal. Bilder and Loughin (1997) pool information across kickers to determine the key factors affecting the success of a field goal. They find that yardage is most important, but that the score at the time of the kick matters, with field goals causing a change in lead more likely to be missed than others. This effect is akin to the clutch (or choke) factor so heavily researched in baseball (see Chapter 2).

For a single team, an interesting set of questions concerns the effective use of that team's kicker. Irving and Smith (1976) build a detailed model of the probability of a successful

field goal for a single kicker. The result of their analysis, a plot showing the probability of a successful kick from any point on the field was used by the coaching staff at the University of California, Los Angeles during the 1972-1973 season to assist in decision-making.

2.3   Evaluating quarterbacks

The quarterback is the player in charge of the offense, the part of the team responsible for trying to score points. He is certainly the most visible player and many argue he is the most critical player on the team. The main skill on which quarterbacks are evaluated is their ability to throw the ball to a receiver for a completed pass. A ball that is thrown and not caught is an incomplete pass and gains no yards. Even worse, a ball that is thrown and caught by an opponent is an interception and results in the opponent taking possession of the ball. The official National Football League system for rating quarterbacks awards points for each completed pass, intercepted pass (negative value), touchdown pass, and yard earned. Essentially the system credits quarterbacks for their passing yardage and includes a 20-yard bonus for each completed pass, an 80-yard bonus per touchdown pass, and a 100-yard penalty per interception. The system has been heavily criticized for favoring conservative short-passing quarterbacks – after all, two 5-yard completions are better rewarded than one 10-yard completion and one incomplete pass. CPT (recall that's Carroll, Palmer and Thorn (1988)) describe the existing system and propose a modest revision of similar form. CPT suggest that there should be no reward for completing a pass and that the touchdown and interception bonuses are too large. Their system appears to be a bit of an improvement, but still does not tie quarterback performance ratings to the success of the offense in scoring points or the success of the team in winning games.

2.4   Summary

Here we have briefly reviewed some of the difficulties with evaluating players, with a focus on kickers and quarterbacks. In this era of greater freedom in player movement from

team to team, research regarding the value of a player or the relative value of two players will become even more crucial. Some of the problems associated with existing methods could be improved by careful application of fairly basic statistical ideas, e.g., examining the proportion of successful kicks rather than the total, examining the yardage contribution of receivers rather than just the number of catches, or considering the yards gained per attempt by running backs.

Two problems associated with evaluating players are more substantial and will be difficult to overcome. These are candidates for future research work. First is the problem of partitioning credit for a play among the various players contributing, e.g., the quarterback and receiver on a pass play. This might be resolved by more detailed record keeping, perhaps an assessment of how much yardage would have been obtained with an "ordinary" receiver. Even then it is difficult to imagine how the contribution of the linemen might be incorporated. One possibility is a plus/minus system like that used in hockey (see Chapter 7) that rewards players on the field when positive events occur (points are scored) and penalizes players on the field when negative events occur (the ball is turned over to the opponent). The second problem with player evaluation is that the focus on yardage gained, although natural, means that more important concerns such as points scored and games won are not used explicitly in player evaluation. For example, all interceptions are treated the same, even those that occur on last-second desperation throws. As part of our discussion of football strategy in the next section, we build up some tools that might be used to improve player evaluation.

3   Football strategy

3.1   Different types of strategy questions

As described in the introduction, professional and college football teams use data on opposing teams' tendencies to prepare for upcoming games. The data have generally not been

used to address a number of other strategy questions that require more statistical thinking. Here we provide examples of some of these types of questions. The first issue concerns point-after-touchdown strategy, football teams have the option of attempting a near certain one-point conversion after each touchdown (probability of success is approximately 0.96) or attempting a riskier two-point conversion (probability of success appears to be roughly $0.40-0.50$). The choice will clearly depend on the score, especially late in the game. Porter (1967) constructs decision rules for end-of-game extra-point strategy, but his method does not make any suggestions for decisions earlier in the game. Another example of a strategy question concerns fourth-down decision-making. Teams have four downs to gain ten yards or must give the football over to their opponents. On fourth down, a team must choose whether to try for a first down with the risk of giving the ball to its opponent in good scoring position, attempt a field goal (worth three points) or punt the ball to its opponent so that the opponent's position on the field is not quite so good. Choosing between the two options clearly depends on the current game situation and also requires reasonably accurate information about the value of having the ball at various points on the field. A key feature of both the point-after-touchdown and fourth-down strategy questions is that the optimal strategy will almost certainly depend on the game situation as measured by the current score and time remaining.

Other strategy questions can be isolated from the game context in the sense that the optimal strategy does not depend on the game situation. For example, Brimberg, Hurley and Johnson (1998) have analyzed the placement of punt-returners in Canadian football (the wider and longer field and more frequent punts make this a more important issue in Canada than it is in the U.S.). They find that a single returner can perform nearly as well as two returners, and that if two returners are used they should be configured vertically, rather than the more traditional horizontal placement (i.e., at the same yard line).

In the remainder of this section we focus on strategy questions that need to be addressed

9

in the context of the game situation (score, time remaining, etc.). We consider two ways of assessing the current game situation and use them to develop appropriate strategies. First, we measure the value of having the ball at a particular point on the field by estimating the expected number of points that a team will earn for a possession starting from the given point. Decisions can then be made to maximize the expected number of points obtained by the team. Following that, we consider a more ambitious proposal, measuring the probability of winning the game from any situation. Strategies may then be developed that directly maximize the probability of winning (the global objective) rather than maximizing the number of points scored in the short-term (a local objective). Making decisions to maximize the team's probability of winning the game would seem to be a superior approach, but we will see that it turns out to be quite difficult to put this idea into practice.

3.2   Expected points

3.2.1   Estimating the expected number of points for a given field position

Carter and Machol (1971) use a data-based approach to estimate the expected number of points earned for a team gaining possession of the ball at a given point. Let $E(\text{pts}|Y)$ denote the expected number of points for a team beginning a series (first down) with the ball Y yards from the opposing team's goal. The natural statistical approach for evaluating the expected number of points is to examine all possible outcomes of a possession starting from the given point, recording the value (in points) of each outcome to the team with the ball and the probability that each outcome will occur. In football there are 103 possible outcomes, four of which involve points being scored: touchdown (7 points ignoring for the moment questions about point-after-touchdown strategy), field goal (3 points), safety (-2 points, i.e., 2 points for the opponent), and opponent's touchdown (-7 points, i.e., 7 points for the opponent). The remaining 99 outcomes cover the cases when the ball is turned over to the opposing team with the opponent needing Z yards for a touchdown, with Z

ranging from 1 to 99. The opponent can expect to score $E(\text{pts}|Z)$ points after receiving the ball Z yards from its target and hence the value of this outcome for the team currently in possession of the ball is $-E(\text{pts}|Z)$.

There are two complications that must be addressed before the expected point values can be determined. First, the probability of each of the outcomes is unknown and must be estimated. Carter and Machol find the probability of each outcome based on data from 2852 series (2852 sequences of plays that began with 1st down and 10 yards to go). The second complication is that, as derived above, the expected number of points for a team Y yards from the goal depends on the expected number of points if its opponent takes possession Z yards from the goal. In total, it turns out that there are 99 unknown values, $E(\text{pts}|Y)$, and these can be found using the 99 equations that define the expected values.

In fact, Carter and Machol chose not to solve this large system of equations with their limited data. Instead, they combined all of the series that began in the same 10-yard section of the the field (e.g., 31-40 yards to go for a touchdown). Their results are provided in Figure 1. The results can be summarized by noting that it is worth about 2 points on average to start a series at midfield (50 yards from the goal line), and every 14 yards gained (lost) corresponds roughly to a 1 point gain (loss) in expected value. Following this rule of thumb, we find that having the ball near the opposing team's goal is worth a bit less than 7 points since the touchdown is not guaranteed, and having the ball near one's own goal is worth a bit more than -2 points (the value of being tackled behind one's own goal). Interestingly, it appears that starting with the ball just beyond one's own 20-yard-line (80 yards from the goal) is a neutral position (with zero expected points) and that is fairly close to the typical starting point of each game. The Carter and Machol analysis was carried out using data from the 1969 season. Football rules have been modified over time and it is natural to wonder about the effect of such rule changes. CPT redid the Carter and Machol analysis using 1986 data and obtained similar results.

11

FIGURE 1 about here

3.2.2 Applying the table of expected point values

It is possible to use the expected point values of Figure 1 to address some of the football strategy issues raised earlier. First, we describe Carter and Machol's use of their results to evaluate the football wisdom that says turnovers (losing the ball to your opponent by making a gross error) near one's own goal are more costly than turnovers elsewhere on the field. From the data in Figure 1 one can see that a turnover at one's own 15-yard-line (85 yards from the target goal) changes a team from having expected value -0.64 to -4.57 (the opponent's value after taking possession is 4.57), a drop of 3.93 expected points. The same turnover at the opponent's 45-yard-line changes the expected points from 2.39 to -1.54, a drop of 3.93 expected points! Turnovers are worth about 4 points and this value doesn't seem to depend on the location at which the turnover occurs.

Next we consider the question of appropriate fourth down strategy. Here is a specific example, consider a team at its opponent's 25-yard-line (25 yards from the goal) with fourth down and one yard required for a first down that will allow the team to maintain possession of the ball. Suppose the offensive team tries to gain the short distance required, then they will either have a first down still in the neighborhood of the 25-yard-line (expected value 3.68 points from Figure 1) or the other team will have the ball 75 yards from their target goal (expected value to the team currently in possession is -0.24 points). Professional teams are successful on fourth-down plays requiring one yard about 70 percent of the time which means that they can expect $(0.7)(3.68) + (0.3)(-.24) \approx 2.5$ points on average if they try for the first down. The result is recorded as an approximation – it ignores the possibility that the offensive team will gain more than a single yard but it also ignores the possibility that the team will lose ground if it turns the ball over. Field goals (kicks worth 3 points if successful) from this point on the field are successful about 65% of the

12

time. Should the field goal miss the other team would take over with 68 yards-to-go under current rules (expected value to the offensive team of approximately -0.65 points). Trying a field goal yields $(0.65)(3) + (0.35)(-.65) = 1.7$ points on average. Clearly teams should go for the first down rather than try a field goal as long as these probabilities are reasonably accurate. In fact, for the specified field goal success rate, we find the the probability of success required to make the fourth-down play the preferred option is about 0.50. CPT investigate a number of such scenarios and find that field goals are rarely the correct choice for fourth-down situations with six or fewer yards required for a first down (at least when evaluated in terms of expected points).

Expected points might also be used to evaluate teams' performances. Teams could, for example, be judged by how they perform relative to expectation by recording the expected number of points and the actual points earned for each possession. If the offensive team begins at their 25-yard-line (75 yards from from the goal) and scores a field goal then they have earned 3 points, 2.76 more than might have been expected at the start of the possession. The contributions of the offense, defense, and special teams (punting and kicking) could be measured separately. It is more difficult to see how this can be applied to evaluating individual players since expected point values are only determined for the start of a possession. Our fourth-down example above indicates how we can assign values to plays other than at the start of a possession, but it becomes quite complicated when we try to assign point values to second- or third-down situations. If expected point values were available for every game situation, then it might be possible to give a player credit for the changes in the team's expected points that result from his contributions. Of course, partitioning credit among the several players involved in each play remains a problem.

3.2.3   Limitations of this approach

There are limitations associated with using expected point values to make strategy decisions. The first limitation is that the Carter and Machol (and CPT) expect point values are based on aggregate data from the entire league. Individual teams might have difference expected values. For example, a team that prefers to advance the ball by running might have lower expected values from a given point on the field than a team that prefers to advance the ball by throwing. A second limitation is that, even if we accept a common set of expected point values, applying the expected point values to determine appropriate strategies requires assessing the probabilities of many different events. For example, if a team's kicker has an extremely high probability of success, or a team has an ineffective offense that is not likely to succeed on fourth down, then the fourth-down strategy evaluation we considered earlier might turn out differently. It is probably not appropriate to think of this as a problem, it merely points out that proper use of Figure 1 requires the user to make determinations of the relevant probabilities. A final limitation is that the expected points approach completely ignores two key elements of the game situation, the score and the time remaining. The correct strategy in a given situation should surely be allowed to depend on these important factors. As an extreme example, consider a team that trails by 2 points with 1 minute remaining in the game and faces fourth down on the opponent's 25-yard-line with one yard needed for a first down. An analysis based on expected points that suggests a team should try for the first down is clearly invalid because at that late point in the game, it is more important to maximize the probability of winning (achieved by trying to kick the field goal) than to maximize the expected number of points. We next consider an approach that treats maximizing the probability of winning as the objective and tries to take into account all of the important elements of the game situation.

3.3 The probability of winning the game

3.3.1 Estimating the probability of winning for a given situation

Carter and Machol's expected number of points for different field positions can be used to make optimal short-term decisions when the time remaining is not a critical element (time remaining is important near the end of the first half and the end of the game). Decisions late in the game need to be motivated more by concerns about winning the game than about maximizing the expected number of points. This motivates an alternative approach to football strategy that requires estimating the probability of winning the game from any current situation. For purposes of this chapter, we define the current game situation in terms of: the current difference in scores (ranging perhaps from -30 to 30), the time remaining (perhaps taking the 60-minute game to consist of 240 15-second intervals), position on the field (1 to 99 yards from the goal), down (1 to 4), and yards needed for a first down that will allow the team to maintain possession (ranging perhaps from 1 to 20). The win probabilities can be estimated for each game situation in a number of different ways. We describe two basic approaches: an empirical approach similar to that used by Carter and Machol, and an approach based on constructing a probability model for football games. Either approach must deal with the enormous number of possible situations. Using the values given above, there are more than 100 million possible situations.

3.3.1.1 An empirical approach

Conceptually at least, we can proceed exactly as Carter and Machol did and obtain probability estimates directly from play-by-play data. For any game situation, we need only record the frequency with which it occurs and the ultimate outcome (win/loss) in each games where the situation occurred. The number of possible situations is far too large for this approach to be feasible. After all, there are more than 100 million situations and only 240 National Football League games per season with 130 plays per game. CPT perform

15

an analysis of this type by restricting attention to the beginning of a team's possession (situations with first down and 10 yards to go), taking the current difference in scores to be between −14 and 14, and taking the time remaining to consist of 20 three-minute intervals. These modifications reduce the number of situations to a more manageable number, $29 \times 20 \times 99 = 57420$. They use two seasons' data to obtain estimates of the probability of winning the game for each of the 57420 situations. For example, the probability that a team beginning the game with first down at its own 20-yard-line ultimately wins the game is .493 according to CPT. By way of comparison, a team starting with first down at its own 20-yard-line but trailing by 7 points with 51 minutes remaining in the game has probability of winning equal to .281. Unfortunately, there is no description of how the win probabilities were actually estimated from the data so it is difficult to endorse them completely. More important, it is not possible to obtain win probabilities for situations that are not explicitly mentioned in the book.

In order to pursue this approach further, we approximate the win probability function derived by CPT using some simple statistical modeling. Using 76 win probability values provided in the book, we derive the following fairly simple logistic approximation that gives the probability of winning, p, in terms of the current score difference, s, the time remaining (in minutes), t, and the yardage to the opposing team's goal, y, at the beginning of a team's possession:

$$\ln\left(\frac{p}{1-p}\right) = .060s \; + \; .084\frac{s}{\sqrt{t/60}} - .0073(y-74),$$

where ln is the natural logarithm. This approximation is motivated by the Stern (1994) model that relates the current score and time remaining to the probability of winning a basketball or baseball game. Note that the logistic equation empirically establishes a team's own 26-yard-line ($y = 74$ yards from the goal) as neutral field position at the start of a possession. For the two situations described in the preceding paragraph this approximation

16

Score	Yards	Time	Estimated win probability		
difference	from goal	remaining	CPT	Logistic	Dynamic programming
0	80	60.0	.493	.489	—
0	80	23.6	.490	.489	—
−7	80	47.4	.274	.246	—
−7	67	21.0	.218	.205	—
−7	74	13.5	.153	.161	—
−8	94	10.5	.025	.097	.097
5	67	7.7	.842	.820	(.811,.816)
−5	74	6.6	.178	.174	(.267,.270)
5	78	2.4	.945	.915	(.981,.989)
−5	58	1.8	.069	.071	(.037,.152)
5	50	1.3	.990	.964	(.998,1.00)

Table 1: Win probabilities from Carroll, Palmer and Thorn (1988), along with two alternatives described in the text.

gives .489 and .249 (compared to the CPT values, .493 and .281). Table 1 compares the values obtained by CPT and those obtained by the logistic approximation for a number of situations. (The final column of Table 1 will be discussed later.)

The probability of winning is shown graphically in Figures 2a-c for selected values of the score difference, s, time remaining, t, and yards from the goal, y. Figure 2a shows the importance of the time remaining. Even relatively modest score differences become significant as the time remaining decreases towards zero. Figure 2b indicates that for the logistic approximation the effect of field position is (for the most part) independent of the score difference and time remaining. Figure 2c shows once again the effect of time remaining with the curve corresponding to less time remaining steeper near zero score difference. Figure 2c also illustrates a weakness of the logistic approximation. Because it is not derived expressly for football, the logistic approximation does not account for the 3- and 7-point scoring increments, i.e., the curves in Figure 2c are continuous. The logistic approximation treats the difference between a 4-point deficit and a 6-point deficit as being no different than the difference between a 7-point deficit and a 9-point deficit. Clearly the latter difference is much more significant, since the 9-point deficit will require the team that

17

is behind to get at least two scores to tie or win whereas the 7-point deficit can be made up with a single score. By contrast, both a 4-point and a 6-point deficit can be overcome by a single score. A second weakness of the logistic approximation is that when the score difference is equal to zero the probability of winning does not depend on the time remaining (this is also visible in Figure 2c as curves with different amounts of time remaining intersect when the score difference is zero). It seems likely that the probability of winning would be higher for a team with $s = 0$, $y = 1$, $t = 1$ (excellent field position near the end of a tie game) than for a team with $s = 0, y = 1, t = 59$ (excellent field position very early in a tie game). Before using the win probabilities to find answers to our strategy questions, we consider another approach to estimating the win probabilities.

FIGURE 2 about here

3.3.1.2   Dynamic programming

Dynamic programming is a technique that can be used to find optimal strategies and simultaneously derive the probability of winning from a given situation under optimal play. We first describe a decision-theoretic formulation of football that allows us to apply dynamic programming. Let's take the two teams in the game to be Team A and Team B. As before, we consider the current situation or state (as it is generally called in dynamic programming) of the football game as being given by: the difference in scores, the time remaining, the position on the field, the down and the yards needed for a first down. In addition we will need to keep track of which of the two teams has possession of the ball so that we add this to the definition of the state. Each state is associated with a value that can be thought of as defining the objective of the game, e.g., we might take the value of a state to be the probability that Team A wins starting in the given state. From any state, the two teams have a limited number of actions from which they must choose. Although there is considerable flexibility in defining this set of actions, for now we restrict attention to the choices

18

available to the team in possession of the ball. Their possible actions include run, short pass, long pass, punt, and field goal. Not every action is reasonable from every state (e.g., we would not try a field goal on first down from our own 5-yard-line), but any reasonable model will avoid choosing these suboptimal actions. Team A should choose the action at each point in the game that will give it the highest probability of winning (i.e., they try to maximize the expected value of the next state) and Team B should choose the action that will give Team A the lowest probability of winning (i.e., they try to minimize the expected value of the next state). We require the distribution of possible outcomes for each of the possible actions (a difficult task that we return to shortly) to solve for the optimal action in a given state. Dynamic programming is an algorithm for finding the optimal action for every state and determining the value of being in that state (this is the probability that Team A wins from that state).

Dynamic programming starts at the end of the game (no time remaining) by defining any state in which Team A is ahead of its opponent as having value one, and any state in which Team A is behind as having value zero. Ties can be given the value one-half. These values, corresponding to the probabilities of Team A's winning the game, are obvious because there is no time remaining in the game. Now, given that we know the value of every state at the end of the game, we can back up one-time unit (15-seconds in the specification used here) and determine optimal strategy for any state with one-time unit remaining. First, we evaluate the expected probability of winning under each action by averaging over the distribution of possible outcomes. Team A should choose the action that gives it the highest expected probability of winning the game. Team B, when it is in possession of the ball, should choose the action that gives Team A the lowest expected probability of winning the game. After determining the optimal strategy and value for every state with one time-unit remaining, we can continue to move backwards from the end of the game. We find the optimal strategies for the states at time t by averaging over the results that we have already

found for future states. Dynamic programming is a powerful computational algorithm for solving complex decision problems like this one.

It remains only to describe how we determine the distribution of outcomes under any action. In theory, it could be obtained by a careful analysis of detailed play-by-play data. Here, a small sample of play-by-play data was used to suggest an approximate distribution. To illustrate, Table 2 gives the distribution for a run play, a short pass play, and a long pass play. Each row of the table gives one possible outcome (yardage gained and an indication of whether the ball has been turned over to the opponent) and the probability that it occurs. These probability distributions were constructed to match known features of the true distributions, e.g., the probability of a lost fumble is .015 and the probability of an intercepted pass is .04 (results cited in CPT and, more recently, in Brimberg and Hurley (1997)). Note that passes may result in an interception or fumble so the probability of a turnover is .055 when averaged over all pass plays. Similarly, the mean gain on a run is just under four yards. The remaining details of the distribution represent a crude estimate based on limited data. The distribution of possible outcomes for punts and field goals were created using a similar procedure. The details of the distributions for these two actions are not provided here.

There is plenty to criticize here, e.g., the use of only a single distribution for all run plays, the use of only two passing distributions (short and long), the discrete approximations to phenomenon that are nearly continuous in nature, the complete exclusion of defensive actions. However, the biggest difficulty with this approach to determining optimal strategy is computational. The state space is enormous and to this point it has only been possible to solve for optimal strategy during the last 10 minutes of the football game. In addition, the strategy findings appear to be quite sensitive to the specified distributions which (in theory) reflect the relative abilities of the two teams. Distributions that are inaccurate may lead to unintended consequences. For example, an earlier version of the distributions in Table 2

Distribution of outcomes for various actions								
Run play			Short pass play			Long pass play		
Yards	Turnover	Probability	Yards	Turnover	Probability	Yards	Turnover	Probability
−4	0	0.020	−5	0	0.030	−10	0	0.045
−2	0	0.060	−5	1	0.010	−10	1	0.005
−1	0	0.065	0	0	0.400	0	0	0.595
−1	1	0.005	3	0	0.065	0	1	0.055
0	0	0.145	5	1	0.025	18	0	0.195
0	1	0.005	6	0	0.140	27	0	0.080
1	0	0.125	8	0	0.130	50	0	0.020
1	1	0.005	8	1	0.010	99	0	0.005
2	0	0.110	12	0	0.075			
3	0	0.090	16	0	0.055			
4	0	0.070	20	0	0.040			
6	0	0.090	35	0	0.017			
8	0	0.060	99	0	0.003			
10	0	0.050						
15	0	0.085						
30	0	0.010						
50	0	0.004						
99	0	0.001						

Table 2: Assumed distribution of outcomes for run plays, short pass plays, and long pass plays. Each play is assumed to consume one 15-second time unit.

led to the conclusion that all teams should always choose to throw long passes (unless ahead and trying to run out the clock). Even with these limitations, the optimal strategies obtained from this model are useful. For one thing they suggest that 2-point conversions after touchdown should be attempted more often than they are in practice. This is based on the current rate of success in United States professional football (approximately 0.50 for the 2-point conversion and 0.96 for the 1-point conversion).

The expected win probabilities produced by the dynamic programming approach are included in Table 1 for comparison with the other methods. Intervals are given when the time remaining is in between two time-units. The dynamic programming results are similar to those obtained by CPT, however, some substantial differences do occur. It appears that the dynamic programming approach allows for a greater probability of come-from-behind wins (likely due to some favorable features of the distribution of outcomes assumed for long passes).

The potential of dynamic programming was realized long ago. The annotated bibliography of the book on sports statistics edited by Ladany and Machol (1977) includes a reference to Casti's (1971) technical report which apparently outlines a similar approach. More recently, Sackrowitz and Sackrowitz (1996) develop a dynamic programming approach to evaluating ball control strategies in football. Their work is similar to that described here except that team possessions are analyzed rather than individual plays. They define a limited set of offensive strategies for a team (ball control, regular play, hurry-up) and assign a distribution for time used by each strategy and a probability of scoring a touchdown for each strategy. Their finding is that a team should not change its style of play for a particular opponent.

Score	Yards	Time	Win probability		
difference	from goal	remaining	Before turnover	After turnover	Decrease
0	25	45	.589	.502	.087
0	50	45	.544	.456	.088
0	85	45	.480	.394	.086
3	25	5	.804	.742	.062
3	50	5	.773	.705	.068
3	85	5	.725	.650	.075

Table 3: Change in win probability due to a turnover for several different scores, field positions, and time remaining.

3.3.2 Applying the estimated win probabilities

We can now return to the types of strategy considerations that were evaluated earlier using expected points. For this discussion, we use the logistic approximation to the win probability (because the CPT results are not available for all of the situations we are interested in). We do not use the dynamic programming results because it is evident that more work is required to make this approach feasible. It should be noted however that the dynamic programming approach is a promising one for addressing detailed strategy questions.

Recall that Carter and Machol (1971) found that the effect of a turnover did not depend on the location on the field where the turnover occurred. It seems likely that the time remaining in the game will make a difference with respect to this issue. Table 3 gives the probability of winning before and after a turnover at several different locations at two different points in the game. Early in the game we find that the Carter and Machol result holds, but later in the game the location of the turnover on the field does matter. Turnovers near your own goal late in a close game are more costly than turnovers near midfield, as intuition might suggest.

Interestingly, the optimal fourth down strategy also depends on the time remaining. Early in the game, win probabilities support the recommendation derived using expected points, teams should go for the first down rather than kick a field goal. However, optimal

late-game strategy appears to be sensitive to the model used for estimating win probabilities. The logistic approximation does not inspire great confidence so we do not provide any details here.

Win probabilities might also be used to evaluate team performances. The offensive part of a football team could, for example, be judged by their net effect on the team's win probability. CPT propose win probabilities for precisely this purpose and work through three games in detail. The CPT approach only estimates win probabilities at the start of each possession so that it would be difficult to use them for evaluating individual plays or players. If win probabilities were available for every possible situation, as they would be if dynamic programming were used to estimate them, then it might be possible to give a player credit for the changes in the team's win probability that result from his contributions. This approach could also be used to assess the effectiveness of running plays and passing plays or the effect of penalties by summing the changes in win probability associated with all plays of a given type. Once again the difficult problem of partitioning credit among the several players involved in each play requires some thought.

3.3.3 Limitations

Conceptually, win probabilities come closest to providing the ideal information needed to make effective strategy decisions. One limitation of this approach is that, as with expected point values, the win probabilities are estimated from aggregate data (using either the CPT or dynamic programming approach) and thus may not be relevant for a particular team or game. The win probability for Team A in a particular situation may be different than if Team B were in the same situation. It still seems that a set of "average" win probabilities would be a useful decision-making tool.

A more important issue at this point in time is the difficulty in obtaining credible estimates for the win probabilities. There are problems with both the empirical approach

of CPT and the dynamic programming approach that we considered. Large amounts of data are required to apply the empirical approach of CPT and to expand the number of situations for which win probabilities are defined. We must also decide how many different situations to address. For example, in professional football in the United States the home team is usually thought to have a three-point advantage, or put another way, the home team wins approximately 59% of all games. Should we compute separate win probabilities for the home and visiting team for each state? Dynamic programming, our second approach to estimating win probabilities has great potential but also requires additional data. Data are needed to construct realistic distributions for the various plays/actions. In addition, it would be good to expand the model to include both offensive and defensive choices of actions at each state. This would make things more realistic than the offense-only model considered here. During games, teams try to outguess each other, so that the offense will try to use a run play when the defense expects a pass play. Incorporating offensive and defensive actions would require the distribution of outcomes for each offensive action under a variety of assumptions about the defensive team's strategy. Unfortunately, this would take our fairly large dynamic programming problem and make it even more complex.

Some researchers have worked in the opposite direction, constructing simpler models that can yield informative results on particular questions, e.g., Brimberg and Hurley (1997) describe a simple model of football and use it to assess the effect of turnovers on the probability of winning a game.

4   Rating of teams

Due to the physical nature of football, teams usually play only a single game each week. This limits the number of games per season to between 10 and 20 games (depending on whether we are thinking of United States college football, United States professional football, or Canadian professional football). The seasons are not long enough for each team to play

every other team. Typically teams are organized in leagues or divisions within which all teams play each other once or twice; however these teams will play different schedules outside of the division. Because teams play unbalanced schedules, an unequivocal determination of the best team is not possible. Playoff tournaments are used to determine champions in professional football but not in major United States college football. There are more than 100 college teams competing at the highest level and a unique champion is not determined on the field of play. The performances of the best teams are judged by a poll of coaches or sportswriters to identify a champion. It is natural to ask whether statistical methods can be used to rate teams and identify a champion. Even though professional football uses a playoff tournament to identify a champion, there is some interest in rating teams there as well, especially in the middle of the season. This is primarily because the question of how to find suitable ratings for teams is closely related to questions concerning prediction of game outcomes and preparation of a betting line. Prediction is covered in Chapter 12 later in the book so here we limit ourselves to a brief review of the work that has been carried out concerning the rating or ranking of football teams.

There has been interest in rating college football teams with unbalanced schedules for a long time. Dickinson (1941) describes an approach that he used in the 1920s and 1930s which gave teams points for each game they won, with the number of points depending on the quality of the opponent. This is an example of a rating method that relies only on a record of which teams have defeated which other teams (with no use made of the game scores). Other examples of this type in the statistical literature include the methods of Bradley and Terry (1952) or Andrews and David (1990) for data consisting of contests/comparisons of two objects at a time. The National Collegiate Athletic Association (NCAA) is the governing body for college sports in the United States and is responsible for determining champions in a variety of sports. The NCAA relies on a measure of this type, the Ratings Percentage Index (a combination of a team's winning percentage, the average of its opponents' winning

26

percentages and the average of its opponents' opponents' winning percentage) in a variety of sports but not football.

An extremely popular approach to rating teams makes use of the scores accumulated by each team during their games. Such ratings have become increasingly popular due to their relevance for prediction (see also Chapter 12). Most often these ratings approaches apply the method of least squares or related normal distribution theory to obtain ratings that minimize prediction errors (Leake, 1976; Stefani, 1977, 1980; Harville, 1977, 1980; Stern, 1995; Glickman and Stern, 1998). We briefly describe the basic idea of these approaches. Suppose that R_i is used to represent the rating for team i and R_j is the rating for team j. When team i plays team j the ratings would predict the outcome as $R_i - R_j \pm H$ where H is a home-field advantage measure (approximately 3 points in professional and college football in the United States) and the sign of H depends on the site of the game. If we use Y to represent the actual outcome when these teams play, then the prediction error is $Y - (R_i - R_j \pm H)$. Given the results from a collection of games, we can estimate the ratings to be those values that make the prediction errors as small as possible, e.g., least-squares ratings minimize $\sum(Y - (R_i - R_j \pm H))^2$. Ratings of this type appear in the *USA Today* newspaper during the college football season. Of course, it is not necessarily true that methods based on normal distributions are appropriate for analyzing football scores. Mosteller (1979) presents a "resistant" analysis of professional scores to prevent unusual scores (outliers) from having a large effect. Bassett (1997) introduces the possibility of using least absolute values in place of least squares in order to minimize the effect of unusual observations. Rosner (1976) builds a model for rating teams or predicting outcomes that makes explicit use of the multiple ways of scoring points in football. Mosteller (1970) and Pollard (1973) provide exploratory analyses of football scores but do not focus on rating team performance.

5 Some other topics

Any presentation of the relationship of probability and statistics to football (or any other sport for that matter) will focus on those aspects of the sport that the author finds most interesting and promising. This section provides references to other work not discussed in detail. We also mention some problems that have not received much attention but might benefit from statistical analysis.

Professional football teams are constructed primarily by two means, teams draft players from college football teams and teams sign "free agents" (players currently without a contract). Evaluating the contributions of players and placing an economic value on those contributions are obviously relevant to making personnel decisions. These issues have not yet received much attention. The player dispersal draft that allocates new players to teams has been around a long time but has also not received much attention. Price and Rao (1976) build a model for evaluating a variety of different player allocation rules. Other business and economic issues are addressed by Noll (1974). In one chapter of that edited volume, Noll carries out an analysis of attendance in many sports including football.

One strategy issue that is not appropriately addressed by any of the discussion here is the effective use of timeouts and other time management strategies. Carter and Machol (1971) discuss this issue briefly in their work on expected points. CPT also discuss the use of timeouts but both discussions are mainly qualitative. As regards time management strategies, Sackrowitz and Sackrowitz (1996) carry out an investigation of time management by asking whether altering one's strategy to use more/less time can increase the probability of winning.

6 Summary

Football teams have expressed a willingness use statistical methods to learn from available data. Most teams keep detailed records of opponents' tendencies and use that information to plan strategy for upcoming games. In addition, Bud Goode has a long history of consulting for professional teams, identifying the key variables correlated with winning football games and then providing advice on how teams might improve their performance with respect to these variables (see, e.g., Goode, 1978). The discussion here shows that more extensive use of statistical methods in football might provide an opportunity for enhanced player evaluation, and improved decision-making. In this era of greater freedom in player movement from team to team, research regarding the value of a player or the relative values of two players will become even more crucial. With respect to decision-making, the results here suggest that football coaches should attempt fewer field goals (worth 3 points) and instead take more fourth-down risks in pursuit of touchdowns (worth 6,7,or 8 points). More complete results about player evaluation and optimal strategy will require more data and a more substantial research effort.

7 References

Andrews, D. M. and David, H. A. (1990). Nonparametric analysis of unbalanced paired-comparison or ranked data. *Journal of the American Statistical Association*, **85**, 1140-1146.

Bassett, G. W. (1997). Robust sports ratings based on least absolute errors. *The American Statistician*, **51**, 99-105.

Berry, D. A. and Berry, T. D. (1985). The probability of a field goal: rating kickers. *The American Statistician*, **39**, 152-155.

Bilder, C. R. and Loughin, T. M. (1997). It's good! An analysis of the probability of success

for placekicks. Technical report, Department of Statistics, Kansas State University, Manhattan, KS, submitted to *Chance*.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324-345.

Brimberg, J., and Hurley, W. J. (1997). The turnover puzzle in American football. Technical report, Royal Military College of Canada, Kingston, Ontario, Canada.

Brimberg, J., Hurley, W. J., and Johnson, R. E. (1998). A punt returner location problem. To appear in *Operations Research*.

Carroll, B., Palmer, P., and Thorn, J. (1988). *The Hidden Game of Football*. New York: Warner Books.

Carter, V. and Machol, R. E. (1971). Operations research on football. *Operations Research*, **19**, 541-545.

Casti, J. (1971). Optimal football play selections and dynamic programming: a framework for speculation. Technical report, Project PAR284-001, Systems Control, Inc., Palo Alto, CA.

Dickinson, F. G. (1941). My football ratings — from Grange to Harmon. Omaha, NE: What's What Publishing Co..

Glickman, M. E. and Stern, H. S. (1998). A state-space model for National Football League scores. *Journal of the American Statistical Association*, **93**, 25–35.

Goode, B. (1978). Relevant variables in professional football. *ASA Proceedings of the Social Statistics Section*, 83-86.

Harville, D. (1977). The use of linear model methodology to rate high school or college football teams. *Journal of the American Statistical Association*, **72**, 278–289.

Harville, D. (1980). Predictions for National Football League games via linear-model methodology. *Journal of the American Statistical Association*, **75**, 516–524.

Irving, G. W. and Smith, H. A. (1976). A model of a football field goal kicker. In *Management Science in Sports* (edited by R. E. Machol, S. P. Ladany, and D. G. Morrison), pp. 47-58. New York: North-Holland.

Ladany, S. P. and Machol, R. E. (editors) (1977). *Optimal Strategies in Sports.* New York: North-Holland.

Leake, R. J. (1976). A method for ranking teams: with an application to college football. In *Management Science in Sports* (edited by R. E. Machol, S. P. Ladany, and D. G. Morrison), pp. 27-46. New York: North-Holland.

Morrison, D. G. and Kalwani, M. U. (1993). The best NFL field goal kickers: are they lucky or good? *Chance*, **6**, No. 3, 30-37.

Mosteller, F. (1979). A resistant analysis of 1971 and 1972 professional football. In *Sports, Games, and Play: Social and Psychological Viewpoints* (edited by J. H. Goldstein), pp. 371-399. Hillsdale, NY: Lawrence Erlbaum Associates.

Mosteller, F. (1970). Collegiate football scores, U. S. A.. *Journal of the American Statistical Association*, **65**, 35-48.

Noll, R. G. (editor) (1974). *Government and the Sports Business.* Washington, DC: The Brookings Institute.

Pollard, R. (1973). Collegiate football scores and the negative binomial distribution. *Journal of the American Statistical Association*, **68**, 351-352.

Porter, R. C. (1967). Extra-point strategy in football. *The American Statistician*, **21**, 14-15.

Price, B. and Rao, A. G. (1976). Alternative rules for drafting in professional sports. In *Management Science in Sports* (edited by R. E. Machol, S. P. Ladany, and D. G. Morrison), pp. 79-90. New York: North-Holland.

Purdy, J. G. (1971). Sport and EDP .... It's a new ballgame. *Datamation*, **17**, June 1,

24-33.

Rosner, B. (1976). An analysis of professional football scores. In *Management Science in Sports* (edited by R. E. Machol, S. P. Ladany, and D. G. Morrison), pp. 67-78. New York: North-Holland.

Ryan, F., Francia, A. J., and Strawser, R. H. (1973). Professional football and information systems. *Management Accounting*, **54**, No. 9, 43-47.

Sackrowitz, H. and Sackrowitz, D. (1996). Time management in sports: ball control and other myths. *Chance*, **9**, No. 1, 41-49.

Stefani, R. T.(1977). Football and basketball predictions using least squares. *IEEE Transactions on Systems, Man, and Cybernetics*, **7**, 117–120.

Stefani, R. T.(1980). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics*, **10**, 116–123.

Stern, H. S. (1994). A Brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, **89**, 1128-1134.

Stern, H. S. (1995). Who's number 1 in college football? . . . and how might we decide? *Chance*, **8**, No. 3, 7-14.

Figure 1. Expected points for a team with first down and ten yards to go from various points on the field and the associated least squares line. Data are from Carter and Machol (1971).

Figure 2. Probability of winning as a function of the score difference, s, the time remaining (in minutes), t, and yards from the goal, y, using the logistic approximation: (a) probability as a function of time remaining for three selected score/yards-from-goal combinations; (b) probability as a function of yards from goal for three selected score/time-remaining combinations; (c) probability as a function of score difference for two time-remaining/yards-from-goal combinations.