SOCCER/WORLD FOOTBALL

Soccer/world football is a sport involving two teams of players kicking and heading a medium-sized ball about on a large rectangular field. Each team has a goal that is quite wide and high and seeks to shoot the ball into the other team's goal. The basic objective is to see which team of the two can score the most goals and thereby win the game. The sport is subject to Laws set down by the Federation Internationale de Football (FIFA). Typically there is a group of teams playing against each other in a league or tournament with a formal schedule of games. A team's intention then is to win a tournament or championship by obtaining the most points, where 3 are awarded for a win, 1 for a tie and 0 for a loss.

The sport of soccer has a long and full history possibly going back to 2500 BC, Henshaw (1979). In the present time the teams each have eleven players. One, the goalie, may handle the ball but only within a restricted area in front of his goal. There is a neutral referee whose decisions are final. The grandest tournament, the World Cup, moves amongst different countries and occurs nominally every four years. The past world champions are Uruguay, Argentina, Brazil, England, France, Germany and Italy. The sport has become a multi-billion dollar business.

Tactics are basic to the games with many styles and plating formations available. A team's choice has been based, in part, on studies of match statistics. The pioneer in the field of match performance analysis is Charles Reep, see Larsen (2001) and Pollard (2002).

1. DATA COLLECTION AND DESCRIPTIVE ANALYSES

Data collection and analysis have been basic to soccer for many years. The most common statistics are a game's result and the number of goals scored by each team. Outputs of data analyses include: reports, tables, graphs, talks, images, and videos.

Many things go on during a game. A basic question is what is to be recorded for analysis and how to do so. Charles Reep, see Larsen (2001) and Pollard (2002), developed solutions starting in the early 1950s. The study of basic events in a game has been called both match performance analysis and match analysis, see Reilly and Thomas (1976), Larsen (2001), McGarry and Franks (2003), and Hughes (2003).

A recent addition to the type of information available is near continuous high-frequency digital spatial-temporal data for individual games. Companies come to a stadium and set up an array of video cameras. By signal processing they then develop the spatial-temporal coordinates of the changing locations of the players on the field, the ball and the referee. Companies doing this include: Match Analysis, Prozone Holdings, and Sport-Universal SA. Di Salvo et al. (2006) have validated one such system. Basic quantities that may now be tabled directly include counts of: tackles, crosses, distances covered, key moments and actions, possessions, passes, interceptions, runs with the ball, incorrect referee decisions, fouls, penalty kicks, challenges, entries into the opponent's...
area, ball touches, blocks, forward passes, long balls, high intensity running, ball velocity and accuracy, balls received. Individual player's performances may be evaluated directly. The data can lead to changes in tactics during a game, see Carling, Williams and Reilly (2009). Substantial soccer databases have been available for many years. Websites include: www.eufa.com, www.fifa.com, www.soccerbase.com, www.soccerway.com, and www.soccerpunter.com.

One early investigation of basic data is that of Moroney (1951), pp. 101-103. He studied the numbers of goals scored in 480 games and prepared a histogram of the counts. In another early study the numbers of successful passes in passing movements were studied for the English First Division during the years 1957-58 and 1961-62 in Reep and Benjamin (1968). The results were presented in tabular form. Reep and Benjamin found a stable correspondence between shots on goal and goals scored of about 10 to 1. Distances traveled by individual players during a game are graphed by player position in Reilly and Thomas (1976). Hughes and Franks (2008) present scatter and time series plots. One early discovery was the existence of a home advantage. Schedules are often set up so that two teams meet both at home and away to deal with this. Jochems (1958, 1962), while addressing the question of whether the Dutch football pools were breaking the gambling laws, found the percentages to be 46.6 for a home win, 31.0 for a visitor win and 22.4 for a draw/tie. Managers make use such data in their decision-making, more so every year. Stochastic models are important in this connection.

2. STOCHASTIC MODELLING

Stochastic models are pertinent to soccer statistics because is there is much uncertainty in what happens and what may happen. During the preceding fifty years stochastic models have been constructed to address soccer questions. Some models may be distinguished as to concerning goals, win-tie-loss, or points. Specific distributions and models that have been employed include: bivariate Poisson, exponential, extreme value, GARCH, generalized linear, logistic, Markov, negative binomial, ordinal, regression, prior, point process, and state space. Many analyses are based on the Poisson distribution whose probability function is given by

\[ \text{Prob}(Y=y) = \mu^y e^{-\mu} / y! \text{ for } y = 0,1,2,... \]

where \( Y \) might represent the number of goals a team scores in a future game. In early work Moroney (1951), using the data mentioned above, compared the number of goals scored by a team in a game to their estimated expected frequency assuming that a Poisson distribution held. On examining the result he was led to seek to improve it and fit a negative binomial distribution. This is a generalization of the Poisson. This fit was satisfactory. However, Greenough et al (2002) found that when data from 169 countries were pooled extreme value distributions were needed.

Reep et al (1968, 1971) work with the numbers of passes in successful passing movements. They fit the negative binomial and found it was good when 0-length cases were excluded.

Suppose next that team i at home is playing against team j visiting. Denote the final score by \((X_{ij}, Y_{ij})\). Various authors have assumed that \(X_{ij}\) and \(Y_{ij}\) are independent Poissons with respective means,

\[ \alpha_i \beta_j \quad \gamma_i \delta_j \]

Maher (1982)
exp\{\alpha + \eta + \gamma_i + \delta_i\}, \exp\{\alpha + \gamma_j + \delta_i\}\) \quad \text{Kuonen (1996) and Lee (1997)}
\end{equation}
\begin{equation}
\exp\{\alpha + \eta + \gamma_i - \delta_i - \phi(e_i - e_j)\}, \exp\{\alpha + \gamma_j - \delta_i + \phi(e_i - e_j)\}\) \quad \text{Rue and Salvesen (2000)}
\end{equation}

Here, \(\eta\) is a home advantage, \(\phi\) is a psychological effect and \(\epsilon_i = \gamma_i + \delta_i\). In their fitting, Dixon and Coles modify the low score model probabilities. The probability that team \(i\) wins the game may be estimated having estimated the model parameters.

Maher (1982) estimates \(\alpha, \beta, \gamma, \delta\) by maximum likelihood and mentions fitting the bivariate Poisson. Karlis and Ntzoufras (2003) give details of fitting a bivariate Poisson studying the data for 24 leagues. They find that the assumption of independence is not rejected in 15 cases out of the 24. They also consider the negative binomial distribution and the inclusion of interaction terms.

The goal difference \(X_{ij} - Y_{ij}\) is also important. It determines whether a game result is win, tie or loss (W, T, or L) and is used to resolve conflicts when the points are equal. Stefani (1980) considers the model

\[X_{ij} - Y_{ij} = \eta + \rho_i - \rho_j + \text{random error}\]

with \(\eta\) the home advantage and the \(\rho_i\) called rankings. Karlis and Ntzoufras (2009) fit the Poisson difference distribution to observed goal differences including a home effect, attacking, and defensive parameters.

Fahrmeir and Tutz (1994) employ a logistic latent variable-cutpoint setup to model an ordinal-valued variable. Their model is

\[
\text{Prob}(Y_{ij} = r) = F(\theta_r + \alpha_i - \alpha_j) - F(\theta_r + \alpha_i - \alpha_j)
\]

for \(r = W, T, L\) with \(F\) the logistic and \(\alpha_i\) the ability of team \(i\). In contrast Brillinger (1996, 2007, 2008, 2009) employs an extreme value variable-cutpoint approach. It leads to the model

\[
\text{Prob}(i \text{ wins at home playing } j) = 1 - \exp\{-\exp\{\beta_i + \gamma_j + \theta_2\}\}
\]

\[
\text{Prob}(i \text{ ties at home against } j) = \exp\{-\exp\{\beta_i + \gamma_j + \theta_2\}\} - \exp\{-\exp\{\beta_i + \gamma_j + \theta_1\}\}
\]

\[
\text{Prob}(i \text{ loses at home against } j) = \exp\{-\exp\{\beta_i + \gamma_j + \theta_1\}\}
\]

with \(\theta_2 > \theta_1\) cutpoints and the standardizations \(\Sigma \beta_i, \Sigma \gamma_i = 0\). The parameters \(\beta_i\) and \(\gamma_i\) represent home and away effects of team \(i\). The extreme value distribution employed is longer tailed to the right. The result of the preceding game, and distances between the cities involved, were considered as explanatories. The \(\beta_i - \gamma_i\) can be interpreted as the advantage of team \(i\) when playing at home. These values were also considered in Leeflang and van Praag (1971). These authors model the goal differences as

\[X_{ij} - Y_{ij} = \eta + \alpha_i + \beta_j + \text{random error},\]

calling the \(\alpha_i\) and \(\beta_j\) "forces". Least squares estimation is employed. Including past results did not improve things. Goodard (2005) fits an ordered probit model, that is \(F\) above relates to the normal, with explanatories.

In a linear regression analysis Panaretos (2002) takes the dependent variable to be points obtained in a game and studies the effect of the explanatories goals scored, goals conceded and time of ball possession. The regression model

\[\text{points} = \alpha + \beta \text{log(goals scored/goals conceded)} + \gamma \text{ball possession} + \text{random noise}\]

is fit and a proportion of variance explained of .971 found.

Barnett and Hilditch (1993) study whether the field being artificial turf affects the game result. They find that the home team's advantage was increased. The use of artificial turf was abandoned.
Attention now turns to the case where time or round, t, plays an essential role in the modelling. Jochems (1958, 1962) defines, as a measure of strength of team i at home and j visiting, \( \lambda_{ij} = W_i - W_j \), with \( W_i \) the average points of team i in its preceding games. The prediction of the result for team i is

- i wins if \( \lambda_{ij} > 0 \)
- i wins if \( \lambda_{ij} = 0 \) home advantage
- i loses if \( \lambda_{ij} < 0 \)

Jochem's study had been commissioned by the Netherlands Lottery Commission to see if any skill was involved on the part of tipsters.

The models listed above can be turned into forecasting procedures by simply adding a t to the subscripts of the parameters and estimating them by using the data up to time t. One then uses the schedule of remaining games to develop forecasts. For example Stefani (1980) considers the model

\[ X_{ijt} - Y_{ijt} = \eta_{ijt} + \rho_i - \rho_j + \text{random error} \]

with \( \rho_i \) a rating based on i's past games and \( \eta_{ijt} = \pm \eta \) represents a home advantage. The quantity \( \rho_i \) represents team i's ability. The fit is by conditional least squares.

Dixon and Coles (1997) employ independent Poissons and introduce an exponentially decaying time function to damp out the effects of previous rounds. The Poisson means are \( \alpha_i(t) \beta_j(t) \eta \) and \( \alpha_j(t) \beta_i(t) \) respectively. The variable t is time or round number and \( \eta \) stands for home advantage. Estimation is by maximizing the likelihood.

Fahrmeir and Tutz (1994) work with a latent variable motivated ordinal model,

\[ \text{Prob}\{R_{ijt} = r\} = \Phi(\theta_r - \alpha_i t - \delta_j t) - \Phi(\theta_r - \alpha_i t - \delta_j t) \]

With R the result, \( r = 0,1,2 \) is a win-tie-loss indicator. Further \( \Phi \) is the logistic, and the \( \alpha_i \) are random walks in t. The computations involve a Kalman filter and simulation. Rue and Salvesen (2000) employ a Bayesian dynamic generalised linear model with independent normals having conditional means given the past

\[ \exp\{\alpha + \eta + \gamma_i - \delta_j\} - \exp\{\alpha + \gamma_j - \delta_i\} \]

and \( \exp\{\alpha + \gamma_j - \delta_i\} + \exp\{\alpha + \gamma_j - \delta_i\} \}

to predict next weekend's results. They employ a directed graph to describe the causal structure of the model as a function of time.

Brillinger (2008, 2009) for each round, t, uses previous round results. He fits a latent variable based model and then uses simulation to make projections of the final points and standings of the teams using the knowledge of the remaining games in the schedule. Estimates are then available for future rounds.

Harvey and Fernandes (1989) study the series of goals scored by England against Scotland in a biyearly match. In a state space approach they assume a negative binomial model with mean an exponentially decaying function of past values.

Next consider modeling the progress of a single game. One approach breaks the course of a game into states corresponding to zones in the field. Pollard and Reep (1997) carry out a logistic regression analysis to model the probability of scoring from various positions on the field. In a series of papers Hirotsu and Wright (2002, 2003a, 2003b) Markov chains with a finite number of states are employed. Hirotsu and Wright (2002) use a four state model to determine the optimal timing of substitution and tactical decisions. The expected number of points to be gained from a change in tactics is considered an objective and dynamic programming is employed. The model includes transition rates of scoring and conceding, and rates of gaining and losing possession. Hirotsu and Wright found evidence for replacing a defender with an attacker when losing.
A common problem is to look for change taking place. Croucher (1984) studied the effect of changing the points awarded. The object of this change is to generate more attacking play. Ridder et al (1994) study the effect of a red card in game. They infer that scoring rate does changes after the ejection having a negative effect on the result from the for the team with fewer players. Other papers studying the effect are: Wright and Hirotsu (2003), Vecer et al (2009) and Volf (2009). The latter takes a point process approach. When team i is playing team j the times of goals are modelled by Poisson processes with log rates
\[ v_{ij}(t) = \lambda_i(t) \exp(\alpha_j + \beta z_{ij_1}(t)) \]
and
\[ v_{ji}(t) = \lambda_j(t) \exp(\alpha_i + \beta z_{ji_2}(t)) \]
respectively t being time. The function \( \lambda_i(t) \) is referred to as the attack intensity for team i and \( \alpha_i \) as the defense parameter. The \( z_{ij} \) are the explanatories:
- \( z_{ij_1}(t) = 1 \) when rival player off with red card, 0 otherwise
- \( z_{ij_2}(t) = 1 \) if actual score positive
- \( z_{ij_3}(t) = 1 \) if actual score negative

\( t \) is time, \( \beta \) and \( z_{ij} \) are 3-vectors. The parameters are estimated using data from the 2006 World Cup. The fitted model may be employed to generate simulations. Dixon and Robinson (1998) also takes a point process approach employing a birth process to model the scoring.

The match video data being collected these days may be processed to determine trajectories of players and the ball. Then models can be developed. Xie et al (2004) take a video of a game, break it into segments, for example "play" or "break". They deal with camera pans and zooms using the green grass ratio and motion intensity. Their analysis uses dynamic programming and spatial-temporal hidden Markov processes. Min et al (2008) work on the same problem making use of rule-based reasoning, Bayesian inference, simulation, and expert systems. Palavi (2008) develops a graph-based multiplayer detection and tracking system. In the simplest case of a notable goal scoring movement in one game Brillinger (2007) develops a potential function approach to obtain a stochastic model.

Berument et al (2006) look for evidence of soccer results affecting the Turkish stock market. They do find an effect for one team as well as a day of the week effect.


3. RANKING

Rankings/ratings/seedings are numerical values meant to describe the relative performances of teams in a league. They are used for a variety of specific purposes. One is the preparation of schedules for knock-out tournaments. Another is for the maximization of the probability that the best competitors meet in a tournament's later stages. If there are N teams the collection of ranking values might be the sequence of integers 1 to N. A famous example is the FIFA/Coca-Cola world ranking, (www.fifa.com.worldfootball/ranking/procedure/men.html) where the values derive from a formula involving major games during the previous 4 years, strength of opponent, W-T-
L points and several other items. It orders the teams of all the countries that are members of the FIFA and is updated steadily as more game results become available.

Jochems (1958, 1962) employed the differences of average game scores and compared this to tipsters' predictions. Hill (1974) uses Kendall's tau to compare tipster results with the end of season standings and finds association. Stefani (1980) suggests the model

\[ X_{ij} - Y_{ij} = \eta + \rho_i - \rho_j + \text{random error} \]


There are papers on ratings for other sports for example Leake (1976), Harville (1977), and Stern (1992).

4. TOURNAMENTS AND SCHEDULING

Scheduling is a basic step involved when games need to be organized amongst the members of a group of teams. Two common types of tournament design are the round robin and the knock-out. In the round robin each pair of teams play each other the same number of times, equally at home and away. The champion of the group is the team with the most points at the end of all the games. In the knock-out case, there is a seeding order and the first round is based on it. The tournament progresses with two teams playing in pairs setup by the seeding order and the winner going on to the following round. This continues until only one team is left, the champion.

Sometimes there are numerical-valued objective functions that can be employed in developing schedules. Their forms can include: distance traveled, total expenses, numbers of consecutive home or away games, and referees' time. There are often constraints that need to be taken note of in the scheduling. For example there may be multiple venues, television schedules, desired dates for two teams from the same city.

Analytic tools employed in an optimization include: integer programming, maximization, knapsack algorithms, simulation and stochastic models.

Kuonen (1996/1997) employs a logistic regression model and shows how output can be used to estimate probabilities of interest for a knock-out tournament. Unlike a round robin future opponents for the following rounds are only know probabilistically. Kuonen works with European Cup Winners Cup data. He investigates three methods for calculating seeding coefficients based on the data for 3 preceding years.

Urban and Russell (2003) consider scheduling competitions on multiple venues, venues not associated with any of the participants. Della Croce and Oliveri (2006) present an integer linear programming approach for scheduling the Italian League. They had to deal with cable TV, and with games between teams in the same city.

Scheduling the Chilean League has also posed special problems. As described in Noronha et al (2007) and Duran et al (2009) these were dealt with via using mathematical programming, recognized existing weaknesses, stadiums availability, international requirements, and reducing travel distance.
Objectives may be described probabilistically and then stochastic models are involved. When the goal is to determine a champion, it is natural to ask questions like: what is the probability that the "best" team actually wins the tournament, or how much better is the first team than the second? Appleton (1995) provides definitions and reviews many tournament structures. He employs simulations.

Scarf et al (2009) provide an extensive review including discussion of: tournament metrics, how the design influences outcome uncertainty, the use of simulation, robustness (to teams dropping out), effects of rule changes, and probability model employed. He defines the parameter

$$P_{qR} = \text{Prob}\{\text{team in top 100q pre-tournament rank percentile progresses}\}$$

Here R is round achieved and $0 < q < 1$. To understand the $P_{qR}$ values he graphs them against q.

It is necessary to schedule the referees also. Yavuz et al (2008) provides an extensive review. The topics covered include: leagues in different countries requiring different objective functions and constraints, tension between referees and clubs caused by past incidents, reducing frequent assignment of the same referee to a team's games. In amateur leagues there may be several games the same day for same referee, perhaps at different places even.

5. GAME THEORY

The expression "game theory" has both a general and a technical meaning. The "technical" comes from the classic von Neumann and Morgenstern (1944) setup. The general refers to the others, particularly tactics. Both meanings provide general frameworks with which to study soccer. A difficulty arising in the theories' applications to soccer is that that game is highly dynamic. The Jones-Tratner (1979) book is unusual in that it starts by emphasizing the defensive formations. Wilkinson (1988) is a book devoted to the topic of tactics.

The formation selected would be controlled by the coach. The 2-3-5 was very common for many years. In it there are 2 defensive specialists, 5 attacking and 3 midfielders in between. Newer formations include: 2-3-5, 5-4-1, 4-5-1, and even 4-1-4-1 with an extra row. The formation 4-4-2 is favored by various famous teams.. The 2 attackers can drop back to assist in obtaining the ball. When the team has possession of the ball, the two outside midfielders can move forward to increase pressure on the opponent. Wilson (2008) provides a review and insightful analysis of many of the formations employed in games dating from 1878 to 2008.

Pollard and Reep (1997) seek to quantify the effectiveness of different playing strategies. They break the action of a game down into discrete events, e.g. pass, center, shoot. They define the "yield" as the probability a goal is scored minus the probability one is conceded. This quantity is used to evaluate the expected outcome of a team possession.

The book Wesson (2002) The Science of Soccer has a chapter on “game theory”. It discusses how the strength and deployment of the team moderates the apparent random motion and discusses the fact that low scoring benefits the weaker team.

One of the momentous occasions in a soccer game is the awarding of a penalty shot. There have been a number of formal studies. Greenless et al (2008) study how the penalty
takers' uniform colour and pre-penalty gaze affects the impressions formed by goal-keeper. When penalties take place to break a tie McGarry and Franks (2000) identify an optimal order for a team to take the shots in. Their conclusion is to begin by ranking the players 1 (best) to 11 (worst). Then use the order: 5, 4, 3, 2, 1, and if necessary after that 6, 7, 8, 9, 10, 11. Jordet et al (2007) considers the roles of stress, skill, and fatigue for answering why the English are poor at penalties. He analyzed 200 shots from World Cup and European Championship. In an analysis of many penalties Franks, McGarry and Hanvey (1999) learned that the in 80% of the cases studied if the non-kicking foot points to the left the ball will be shot to the left. They developed a training program to test this discovery.

Larsen (2001) describes the arrival of Reep's ideas in Norway and the success that teams adopting it enjoyed.

The theory of von Neumann and Morgenstern provides general definitions and leads to random strategies. Haigh (2009) provides the following intuitive example. The table shows estimated percentages of successful penalty shots based on 1876 attempts.

<table>
<thead>
<tr>
<th></th>
<th>Goalie Left</th>
<th>Goalie Center</th>
<th>Goalie Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kicker</td>
<td>Left</td>
<td>Center</td>
<td>Right</td>
</tr>
<tr>
<td>Left</td>
<td>60</td>
<td>90</td>
<td>93</td>
</tr>
<tr>
<td>Center</td>
<td>100</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Right</td>
<td>94</td>
<td>85</td>
<td>60</td>
</tr>
</tbody>
</table>

The situation can be considered is a two-person zero-sum game. If the kicker wins by scoring, the goalie loses. If the goalie wins by saving, the kicker loses. Haigh finds that Nash equilibrium theory leads to the goalie's choosing to move left, center or right with the respective probabilities of 0.44, 0.13, 0.43. For the kicker they are 0.37, 0.29, 0.34. If either player uses their strategy 80% of shots would be scored in the long run. This turns out to be close to the actual percentage.

Other references to applications of the formal theory include: Pacacios-Huerta (2003), Hirotsu and Wright (2006), and Coloma (2007). Hirotsu and Wright consider a zero-sum game focusing on who wins the game, while Hirotsu et al (2009) focus on points gained and where one has a non-zero-sum game. Bennett et al (1980) use hypergame analysis involving the ranking of various forms of hooligans' actions and authorities' reactions followed by hooligan's reactions and so on followed by exploring implications of such scenarios.

6. ECONOMICS AND MANAGEMENT

Reilly and Williams (2003) write, "In the 1980's it became apparent that the football (soccer) industry and professionals in the game could no longer rely on the traditional methods of previous decades. ... Methods of management science were applied to organizing the big soccer clubs and the training of players could be formulated on a systematic basis." Desbordes (2007), Chapter 10, advocates the introduction of modern management tools into soccer. This makes sense if for no other reason than the fact that billions of dollars are now involved in the sport. One does need to remember that at the
youth and amateur levels the amount of money involved is little. Really all that is needed for a game is a ball and a field or a beach. This is the charm of the game.

Management refers to a sport's structure, owners, sponsors, marketers, financiers, schedulers, and others with some form of control. Management provides the superstructure. The fans expect both their team’s players and its management to be successful in acquiring players, providing facilities, and, hiring the coach. Management assists by supporting the enterprise generally. It provides the infrastructure. Other things that management is responsible for include cash inflows-outflows, advertising, manpower planning and hiring, planning new and keeping up old stadiums, and signing players. The stadium owners provide facilities for security, ticket collection, food, souvenirs, and parking. Management studies the performance of their players and others. The books Covell at al (2007) and Carling et al (2009) are references.

Sometimes there is a numerical-valued objective function for the management to work with. Then optimization methods may be employed to good advantage. Other concerns are: loss functions, earnings/revenue, prestige/honor, recruitment, salary negotiation, scouting, referees, merchandise, commercialism, productivity, profitability, travel, television rights.


In a study of competitive balance in the Dutch League Koning (2000) employs the values $C_{ij}$, corresponding to W, T, L, generated via

$$D_{ij} = \varphi_i - \varphi_j + e_{ij}$$

with the $e$'s mean 0, common variance normals

$$C_{ij} = \begin{cases} 1 & \text{if } D_{ij} > c_2 \\ 0 & \text{if } D_{ij} = 0 \\ -1 & \text{if } D_{ij} < c_1 \end{cases}$$

It is used to study changes in competitive balance in Holland.

A novel application of programming theory is presented in Mosheiov (1998). He applies an integer program to the generalized knapsack problem to find a 'Dream Team'. Player values were established by a public poll with the participants given a budget of 2.5 million dollars each. A knapsack algorithm was then employed to establish the best team at the lowest price.

Calster et al (2008) studies an unwanted happening in soccer, the scoreless draw. The occurrence is related it to indices of a team's offensive performance including total goals and earned points per game.

7. SOME REMAINING TOPICS

The book Sellin (1976) The Inner Game of Soccer, provides details of the game from a referee's standpoint. It discusses the laws and the mechanics and talks of the
characteristics of the best referees and the ones who can turn benign games into ugly unsportsmanlike confrontations.

The book Wesson (2002) lays out some of the physics of soccer. For example Chapter 4 analyses the movement of a ball spinning in flight and how players can make it bend. Hall (2005) elaborates on that describing the physical forces at work in a double banana shot, which curves in one direction then swerves in another. Such a shot is very difficult for the goalie to handle.


There are many papers on employing stochastic models in evaluating bets. Various of the papers referred to above go on to study the efficiency of various gambling strategies.

What is ahead for the game? Wright (2009) writes on 50 years of operations research in sport to provide a background.. The future is discussed in The Orange Future of Football Report (2008) and by the sports writer P Gardner (2009) in World Soccer. He suggests the following changes: bigger goals, smaller penalty area, a revised offside rule, fouls denoted contact or technical, red carded player punished but replacement allowed. Reviewing the literature and practice suggest that one can expect to see much more in the way of the collection and analysis of the spatial-temporal data collected in a game.

8. EXTENDED REFERENCE LIST


SUP AMISCO PRO http://213.30.139.108/sport-universal/uk/amiscopro.htm


David R. Brillinger

Statistics Department, University of California, Berkeley, California

Keywords: association football, efficiency; FIFA; game theory; laws of the game; match analysis, notation analysis, performance; ranking; sports science; spatial-temporal data; tactics; time series.