Open

Strategy selection and outcome prediction in sport using dynamic learning for stochastic processes

David Frank Percy*

University of Salford, Manchester, UK

Stochastic processes are natural models for the progression of many individual and team sports. Such models have been applied successfully to select strategies and to predict outcomes in the context of games, tournaments and leagues. This information is useful to participants and gamblers, who often need to make decisions while the sports are in progress. In order to apply these models, much of the published research uses parameters estimated from historical data, thereby ignoring the uncertainty of the parameter values and the most relevant information that arises during competition. In this paper, we investigate candidate stochastic processes for familiar sporting applications that include cricket, football and badminton, reviewing existing models and offering some new suggestions. We then consider how to model parameter uncertainty with prior and posterior distributions, how to update these distributions dynamically during competition and how to use these results to make optimal decisions. Finally, we combine these ideas in a case study aimed at predicting the winners of next year's University Boat Race.

1. Introduction

All sports generate serial measurements collected 'within-game' (such as goals, tackles, runs, wickets and lap times that arise during play) and 'between-game' (such as aggregate totals observed at the end of a match or competition), which we can model with various types of stochastic process. The published literature in operational research contains many examples of this, including an early article by McGarry and Franks (1994) and recent papers by Smith (2007) and Dalang *et al* (2014).

Brillinger (2007) and Stern (2009) present interesting operational research articles on the use of regression analysis for modelling within-game activity and subsequent results in football and cricket. Indeed, by treating time as a predictor variable and the measurement of interest as the response variable, regression analysis can be regarded as a tool of automatic learning and is the most natural way for adapting parameters of static models to allow for new longitudinal and cross-sectional data that arise during the course of a match, tournament or season. Regression models are often constructed to enable easy evaluation and testing, with fairly simple practical interpretations. Variations including logistic, log-linear, normal and tobit regression provide sufficient flexibility to deal with discrete, continuous and mixed performance measures.

However, regression models are generic forms that usually treat all past observations with equal importance. This paper focusses on stochastic processes that typically attach more weight to recent data than to historical data. This potentially offers greater accuracy in resolving evolving decision problems dynamically. Nevertheless, regression models can be present implicitly in all the settings considered here, in order to adapt parameters to different (possibly changing) factors.

The general purpose of stochastic process modelling is to enable statistical analyses that generate optimal decisions relating to strategy selection and outcome prediction. These analyses are enhanced by revising decisions in real time as the relevant sports competitions are under way, using the techniques of Bayesian updating or dynamic learning. Such investigations can benefit sporting participants, who might improve their performances by choosing appropriate courses of action during play, and those with gambling interests, who might wish to determine the probabilities of various events as play progresses.

This paper reviews some of the sports, models and analyses that have been published to address these principles and objectives. We explore commonality and disparity among these publications, and consider deficiencies that might be resolved. In particular, our investigations reveal a prevalence of a restricted class of stochastic processes, with little attention given to dynamic learning. We consider the feasibility, practicality and benefits that this aspect might contribute to these and other sports. It is convenient for classification purposes to consider discrete-time and continuous-time stochastic processes separately, though the distinction can be arbitrarily blurred depending upon the time intervals involved. Similarly, performance outcome measures can be classified as discrete, continuous or mixed random variables.

*Correspondence: David Frank Percy, Salford Business School, University of Salford, Manchester M5 4WT, UK.

Discrete-time models are useful for sports such as golf, cricket, snooker and tennis, as they generate sequences of observations corresponding to scores after countable numbers of holes, balls, shots and rallies have been played. We particularly consider discrete-time Markov chains and time-series models as suitable for sports like these. Continuous-time models are useful for sports such as football, hockey, athletics, swimming, speed skating and basketball, as they generate sequences of observations corresponding to goals scored, changes of possession, changes of lead and baskets scored after uncountable periods of time. We particularly consider continuous-time Markov chains and point process models as suitable for such observations in sports like these.

The inferential aspects of this paper relate to dynamic learning in the form of prior–posterior analysis and sequential updating, and decision analysis in the form of outcome prediction and strategy selection. The applications that we consider to demonstrate these modelling techniques and analytical methods include cricket, football and badminton. We finish with a case study designed to illustrate all these methods in the context of predicting who will win next year's University Boat Race.

2. Discrete-time stochastic processes

In this section, we consider random vectors $\mathbf{x}_n$ that are observable at discrete time points $n \in \mathbb{N}$, where the sequence of observations begins at time 0. Component $i = 1, 2, \ldots$ of $\mathbf{x}_n$ is a random variable $X_{i,n}$ that takes one of these forms: discrete with finite support such as $X_{i,n} \in \{1, \ldots, m\}$; discrete with countably infinite support such as $X_{i,n} \in \mathbb{N}$; continuous with uncountable support such as $X_{i,n} \in \mathbb{R}$; mixed (part discrete, part continuous). There are four common approaches to modelling and analysing the multivariate stochastic process $\{\mathbf{x}_n\}$:

- Fit a joint probability distribution to $\mathbf{x}_n$ with parameters that vary over time. Suitable discrete candidates are the multivariate Poisson distribution, the multinomial distribution and the multivariate hypergeometric distribution. See Maher (1982) and Dixon and Coles (1997) for examples of this approach that use 'between-game' observations arising from football games. Continuous options include the multivariate normal distribution, after linear transformation of $\mathbf{x}_n$ if required.

- Adopt a multi-stage procedure by defining sequentially independent stochastic processes $\{X_{1,n}\}$, $\{X_{2,n}|X_{1,n}\}$, etc. Repeatedly applying the multiplication law of probability then constructs a joint probability distribution if required. This approach reduces the multivariate problem to one of modelling and analysing several univariate random variables. It induces a lack of symmetry among the components $\mathbf{x}_n$, which is desirable for some sporting applications and inappropriate for others.

- General constructions of joint distributions can be derived from marginal distributions using copulae as described by

Nelsen (1999) and applied to football by McHale and Scarf (2011). This also has the advantage of reducing a multivariate problem to several univariate problems and has the advantage of retaining symmetry if required.

- Combine the components of $\mathbf{x}_n$ into a single random variable X_n by means of a simple transformation. Although this formulation generates a simple univariate summary of the process at time point n, the specification can be unnatural and inconvenient for subsequent modelling and analysis.

A sporting illustration of these concepts is obtained by reviewing the scores after each over in a game of cricket, which may be categorised as 'within-game' data collection. In this context, the random vector $\mathbf{x}_n$ might be defined with two components that represent the number R_n of runs scored by the batting side and the number W_n of wickets taken by the bowling side after $n \in \mathbb{N}$ overs. Although we could fit a bivariate distribution for R_n and W_n directly from observed data, the natural modelling approach is to analyse the bivariate stochastic process $\{\mathbf{x}_n\}$ as a multi-stage model by defining separate stochastic processes $\{W_n\}$ and $\{R_n|W_n\}$ and applying the multiplication law of probability $p(r_n, w_n) = p(r_n|w_n)p(w_n)$. This is the approach used by Duckworth and Lewis (1998), who famously devised a method for predicting final scores conditional upon numbers of wickets remaining in one-day cricket. It is also possible to apply the fourth approach above to this cricket scenario by defining a combination of the component random variables R_n and W_n as

$$\mathbf{x}_n = \begin{pmatrix} R_n \\ W_n \end{pmatrix} \Rightarrow X_n = 11R_n + W_n + 1$$

for $R_n = 0, 1, 2, \ldots$ and $W_n = 0, 1, \ldots, 10$. With this notation, the possible values for X_n are countably infinite and an observed score of 148 for 7 after 32 overs corresponds to $r_{32} = 148$, $w_{32} = 7$ and $x_{32} = 1636$, for example. However, this formulation does not lend itself well to further analysis.

2.1. Markov chains

A discrete-time Markov chain for random vector $\mathbf{x}_n$ satisfies

$$p(\mathbf{x}_{n+1} \mid \mathbf{x}_n, \ldots, \mathbf{x}_0) = p(\mathbf{x}_{n+1} \mid \mathbf{x}_n) \quad (1)$$

for $n \in \mathbb{N}$. In the case where $\mathbf{x}_n = X_n$, this simplifies to the univariate model

$$p(x_{n+1} \mid x_n, \ldots, x_0) = p(x_{n+1} \mid x_n) \quad (2)$$

and transition probabilities may then be defined by

$$p_{i,j} = P(x_{n+1} = j \mid x_n = i), \quad (3)$$

where $\sum_{j=1}^{\infty} p_{i,j} = 1$ for $i \in \mathbb{N}$. If the support of X_n is finite, we can easily analyse this model numerically using the transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots \\ p_{m,1} & \cdots & p_{m,m} \end{pmatrix} \quad (4)$$

to derive multi-step transition probabilities and steady-state limiting probabilities using the Chapman–Kolmogorov equations. A fundamental discrete-time Markov process is the simple random walk, which is suitable for modelling the progress of games, losses incurred by gamblers, prices of shares and movement of animals. However, the theory extends readily to include other models for discrete outcome measures that are observed in discrete time, such as the Bernoulli process, branching processes and hidden Markov models.

Models based on discrete-time Markov chains are easy to specify, have many real applications and lend themselves well to numerical computation. For these reasons, they are popular in practice and many researchers have adopted this type of stochastic process. Early articles of this nature considered the sports of tennis (Kemeny and Snell, 1960), squash (Wright, 1988) and one-day cricket (Clarke, 1988). Tennis was also the subject of several extended models, including those by Riddle (1988), Sadovskiǐ and Sadovskiǐ (1993) and Spanias and Knottenbelt (2013), and variations specifically aimed at predicting match outcomes using combined player statistics (Barnett and Clarke, 2005) and common-opponent models (Knottenbelt *et al*, 2012). Other sports analysed by means of discrete-time Markov chains include Australian football (Clarke and Norman, 1998), curling (Kostuk and Willoughby, 1999), badminton (Percy, 2009), table tennis (Pfeiffer *et al*, 2010) and golf (Maher, 2013).

2.2. Time series

Introduced by Box and Jenkins (1970), autoregressive moving average (ARMA) models take the form

$$X_n = \sum_{i=1}^{p} \theta_i X_{n-i} + \sum_{j=1}^{q} \phi_j \epsilon_{n-j} + \epsilon_n \tag{5}$$

for $p, q \in \mathbb{N}$, where ϵ_n are random variables (residual components) with $E(\epsilon_n) = 0$, $\text{var}(\epsilon_n) = \sigma^2$ and $\text{cov}(\epsilon_{n_1}, \epsilon_{n_2}) = 0$ for $n_1 \neq n_2$. These models generally make the stronger assumptions of exchangeability and normality $\epsilon_n \sim N(0, \sigma^2)$ and are popular for modelling chronologically ordered sequences of data, as are their vector counterparts

$$\mathbf{x}_n = \sum_{i=1}^{p} \Theta_i \mathbf{x}_{n-i} + \sum_{j=1}^{q} \Phi_j \boldsymbol{\epsilon}_{n-j} + \boldsymbol{\epsilon}_n. \tag{6}$$

They include white noise as a special case and can be modified easily to allow for known serial patterns. Differencing (integrating) $X_n = Y_n - Y_{n-1}$ can be used to remove trends within or between games, perhaps as a footballer becomes tired or a cricketer gains confidence. Seasonal differencing $X_n = Y_n - Y_{n-s}$ can be used to remove seasonal effects within or between games, perhaps according to which tennis player serves or which rugby team has possession.

However, structural time-series models are generally more robust due to their underlying physical justifications. Introduced

by West and Harrison (1989), a general formulation of state space model (dynamic linear model) consists of a measurement equation

$$X_n = \boldsymbol{\alpha}' \boldsymbol{\mu}_n + \epsilon_n \tag{7}$$

and a transition equation

$$\boldsymbol{\mu}_n = \mathbf{B} \boldsymbol{\mu}_{n-1} + \boldsymbol{\eta}_n \tag{8}$$

for $n \in \mathbb{N}$, in terms of a parameter vector $\boldsymbol{\alpha}$, latent state vectors $\boldsymbol{\mu}_n$, residual random variables $\epsilon_n \sim N(0, \sigma^2)$, a parameter matrix $\mathbf{B}$ and residual random vectors $\boldsymbol{\eta}_n \sim Mn(\mathbf{0}, \mathbf{T})$. The vector equivalent is defined by

$$\mathbf{x}_n = \mathbf{A} \boldsymbol{\mu}_n + \boldsymbol{\epsilon}_n \tag{9}$$

and

$$\boldsymbol{\mu}_n = \mathbf{B} \boldsymbol{\mu}_{n-1} + \boldsymbol{\eta}_n \tag{10}$$

for $n \in \mathbb{N}$, in terms of parameter matrices $\mathbf{A}$ and $\mathbf{B}$, latent state vectors $\boldsymbol{\mu}_n$ and residual random vectors $\boldsymbol{\epsilon}_n \sim Mn(\mathbf{0}, \sum)$ and $\boldsymbol{\eta}_n \sim Mn(\mathbf{0}, \mathbf{T})$. State space models include ARIMA models as special cases.

Time-series models offer much flexibility and extend easily to incorporate cross-sectional regression analysis to complement the longitudinal repeated-measures analysis that is their main feature, thus they are proving very popular for sports. Applications of time-series models to football were published by Knorr-Held (2000), Crowder *et al* (2002) and Owen (2011). Glickman and Stern (1998), Stefani (2009) and Percy (2011a) considered American football, rugby and Alpine skiing, respectively, while Urban (2012) and Cattelan *et al* (2013) considered basketball.

Other discrete-time models for continuous outcome measures include the Gaussian random walk, which is characterised by exchangeable, normally distributed increments in discrete time. Defining

$$X_n = \sum_{i=1}^{n} Y_i \tag{11}$$

in terms of mutually independent random variables

$$Y_i \sim N(\mu, \sigma^2) \tag{12}$$

for $i = 1, \ldots, n$, the sampling distribution becomes

$$X_n \sim N(n\mu, n\sigma^2). \tag{13}$$

Possible sporting applications for this model include cumulative times for events that comprise several legs such as Alpine skiing and relay races, and performances on individuals' successive attempts for athletic field events. In each case, simple transformations of the outcome measures onto the set $\mathbb{R}$ of real numbers might be required before analysis.

3. Continuous-time stochastic processes

In this section, we consider random vectors $\mathbf{x}(t)$ that are observable at continuous time points t for $t \in \mathbb{R}^+$. As in Section 2, component $i = 1, 2, \ldots$ of $\mathbf{x}(t)$ is a random variable $X_i(t)$ that takes one of these forms: discrete with finite support such as

$X_i(t) \in \{1, \ldots, m\}$; discrete with countably infinite support such as $X_i(t) \in \mathbb{N}$; continuous with uncountable support such as $X_i(t) \in \mathbb{R}$; mixed (part discrete, part continuous). As for discrete-time stochastic processes, there are four common approaches to modelling and analysing the multivariate stochastic process $\{\mathbf{x}(t)\}$, with the same benefits and drawbacks:

- Fit a joint probability distribution to $\mathbf{x}(t)$ with parameters that vary over time, using standard functional forms.
- Define sequentially independent stochastic processes $\{X_1(t)\}$, $\{X_2(t)|X_1(t)\}$, etc, and repeatedly apply the multiplication law of probability.
- Combine the marginal distributions of $X_i(t)$ to form a joint probability distribution using suitable copulae.
- Combine the components of $\mathbf{x}(t)$ into a single random variable $X(t)$ to summarise the process at time t, by means of a simple transformation.

A sporting illustration of these concepts is obtained by considering the evolving match score during a game of football. In this case, the random vector $\mathbf{x}(t)$ might be defined with two components that represent the 'within-game' numbers of goals $A(t)$ and $B(t)$ scored by teams A and B, respectively, by time t. The natural formulation here is to fit a bivariate probability distribution to $A(t)$ and $B(t)$, which has the advantage of retaining the underlying symmetry of the two random variables. This technique was published in several articles including those by Maher (1982) and Dixon and Coles (1997). Although these papers adopt this bivariate approach for 'between-game' data, the underlying principle is the same as for 'within-game' data. The latter is relatively scarce in the literature, though Dixon and Robinson (1998) considered bivariate birth processes to model evolving 'within-game' match scores.

Nevertheless, we could model and analyse the bivariate stochastic process $\{\mathbf{x}(t)\}$ as a multi-stage model by defining separate stochastic processes $\{A(t)\}$ and $\{B(t)|A(t)\}$ and applying the multiplication law of probability $p\{a(t), b(t)\} = p\{b(t)|a(t)\}p\{a(t)\}$. For example, we might consider the number of goals scored by team A to be Poisson with constant mean and the number of goals scored by team B to be Poisson with mean that depends on the number of goals scored by team A. This approach again reduces the bivariate problem to one of modelling and analysing two univariate stochastic processes but induces undesirable asymmetry as the marginal distribution for team B is not Poisson. Copulae could also be used to avoid this problem, as demonstrated by McHale and Scarf (2011).

Finally, we could instead combine the two components of $\mathbf{x}(t)$ into a single random variable $X(t)$ that summarises the match score at time t,

$$\mathbf{x}(t) = \begin{pmatrix} A(t) \\ B(t) \end{pmatrix} \Rightarrow X(t)$$

$$= \frac{\{A(t) + B(t)\}\{A(t) + B(t) + 1\}}{2} + A(t) + 1$$

by tabulating the possible scores as triangular numbers and using known results for accumulating these numbers. With this notation, the possible values for $X(t)$ are countably infinite and an observed score of $(2, 1)$ after 65 min of play corresponds to $a(65) = 2$, $b(65) = 1$ and $x(65) = 9$, for example. However, such a specification is unnatural and the bivariate formulation is easier to model and analyse.

3.1. Markov chains

A continuous-time Markov chain for random vector $\mathbf{x}(t)$ satisfies

$$p\{\mathbf{x}(t_{n+1}) \mid \mathbf{x}(t_n), \ldots, \mathbf{x}(t_0)\} = p\{\mathbf{x}(t_{n+1}) \mid \mathbf{x}(t_n)\} \quad (14)$$

for strictly increasing sequence (t_n) with $t_n > 0$ and $n \in \mathbb{N}$. The transition probabilities are

$$p_{i,j}(t_{k+1} - t_k) = P\{\mathbf{x}(t_{k+1}) = j \mid \mathbf{x}(t_k) = i\}, \quad (15)$$

where $p_{i,j}(t)$ are the solutions of the ordinary differential equation

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R} \quad (16)$$

with initial condition $\mathbf{P}(0) = \mathbf{I}$. In this equation, $\mathbf{R}$ is a matrix of transition rates such that $\sum_{j=1}^{\infty} r_{i,j} = 0$ for $i = 1, 2, \ldots, m$. Fundamental types of Markov process include birth-and-death processes, the gamma process and the Wiener process, which is characterised by stationary, independent, normally distributed increments in continuous time. This generalises in distribution to the Lévy process, which is used as a basis for advanced research in mathematics, economics and physics.

Although continuous-time Markov chains are often better conceptual and physical models than their discrete-time counterparts, the need to specify and solve systems of ordinary differential equations can deter practitioners from adopting them. Consequently, few examples of this approach to dynamic modelling appear in the literature. A notable exception was published by Hirotsu and Wright (2002), who considered continuous-time strategy selection during play in the context of football.

3.2. Point processes

A non-homogeneous Poisson process $N(t)$ counts events to time $t \in \mathbb{R}^+$ and can be defined by the conditions of initialisation

$$N(0) = 0, \quad (17)$$

independence

$$t > s \Rightarrow \{N(t) - N(s)\} \amalg N(s) \quad (18)$$

and distribution

$$t > s \Rightarrow \{N(t) - N(s)\} \sim Po\{\Lambda(s, t)\}, \quad (19)$$

where

$$\Lambda(s, t) = \int_s^t \lambda(r)dr \quad (20)$$

and $\lambda(t)$ is a specified intensity function. Common forms of intensity function for general application are constant

$$\lambda(t) = a, \tag{21}$$

power-law

$$\lambda(t) = at^b \tag{22}$$

and loglinear

$$\lambda(t) = ab^t. \tag{23}$$

The first of these special forms corresponds to the homogeneous Poisson process, which itself corresponds to an exponential renewal process. This then generalises in distribution to the class of renewal processes, which also have applications in sport. Other extensions include the familiar compound and mixed Poisson processes, and the hybrid intensity models of Percy *et al* (2010). Perhaps due to its widespread popularity, football leads the way in published applications of point processes in sporting contexts; see Volf (2009) for an excellent description. Again, though, discrete-time models dominate over point processes in practice because their implementation is generally considered to be easier.

4. Dynamic learning

All of the models described above can be expressed in terms of sampling distributions that contain unknown parameters representing transition probabilities, time-series coefficients, transition rates or intensity function coefficients. We assume that the random vector $\mathbf{x}$ has conditional probability mass or density function $f(\mathbf{x}|\boldsymbol{\theta})$ given a parameter vector $\boldsymbol{\theta}$ with prior probability density function $g(\boldsymbol{\theta})$.

For discrete-time Markov chain models, the transition probabilities are dependent and Dirichlet distributions are appropriate. For continuous-time Markov chain models, the transition rates are dependent and conditional normal distributions are appropriate. For other models, without specific knowledge to the contrary, we assume prior independence of the parameters in $\boldsymbol{\theta}$. In this case, their joint prior probability density function takes the form

$$g(\boldsymbol{\theta}) = \prod_{i=1}^{q} g_i(\theta_i) \tag{24}$$

and we assume that the marginal prior distributions of these parameters, after linear transformations if required, take the forms

$$\theta_i \in (0, 1) \Rightarrow \theta_i \sim Be(a, b)$$

$$\theta_i \in (0, \infty) \Rightarrow \theta_i \sim Ga(a, b)$$

$$\theta_i \in (-\infty, \infty) \Rightarrow \theta_i \sim No(a, b) \tag{25}$$

for $i = 1, \ldots, q$, as suggested by Percy (2011b).

Now consider a prior–posterior analysis. Whatever the form of prior distribution for the parameter vector $\boldsymbol{\theta}$, we can update it on observing data $\mathbf{x}$ to generate a joint posterior probability density function of the form

$$g(\boldsymbol{\theta} \mid \mathbf{x}) \propto f(\mathbf{x} \mid \boldsymbol{\theta})g(\boldsymbol{\theta}) \tag{26}$$

using the multiplication law of probability. This function contains all available information about $\boldsymbol{\theta}$ given $\mathbf{x}$. As a competition progresses and we observe more data, we could update the posterior distribution iteratively using the hierarchical procedure

$$g(\boldsymbol{\theta} \mid \mathcal{D}_n) \propto \mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_n)g(\boldsymbol{\theta}) \tag{27}$$

for $n \in \mathbb{N}$, where

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_n) = f(\mathbf{x}_1, \ldots, \mathbf{x}_n \mid \boldsymbol{\theta}) \tag{28}$$

is the likelihood function of the parameter vector $\boldsymbol{\theta}$ given the observed data $\mathcal{D}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. However, this updating algorithm can be very inefficient, and thus instead we use the equivalent and efficient, sequential procedure

$$g(\boldsymbol{\theta} \mid \mathcal{D}_n) \propto \mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_n \setminus \mathcal{D}_{n-1})g(\boldsymbol{\theta} \mid \mathcal{D}_{n-1}) \tag{29}$$

for $n \in \mathbb{N}$, where $\mathcal{D}_n \setminus \mathcal{D}_{n-1} = \{\mathbf{x}_n\}$ is the set difference and $\mathcal{D}_0 = \emptyset$ is the empty set. This sequential updating algorithm is particularly useful for interactive dynamic learning while sporting competitions are in progress. Its use for 'within-game' analysis is relatively new, though Glickman and Stern (1998), Knorr-Held (2000), Crowder *et al* (2002) and Owen (2011) considered applications to 'between-game' situations. Furthermore, Congdon (2003) presented an accessible account of Bayesian inference for several dynamic models, including the time-series models that we review and apply here.

An interesting aspect of this updating procedure arises when there is a choice of time interval for re-defining the available set of observed data. Three possibilities are to update: (a) at all specified time points, such as every time a tennis shot is played; (b) at each important event, such as every time a point is scored; (c) whenever a major decision is required, such as every time a game is completed. The choice will depend upon the amount of data available, the robustness of the chosen model and practical requirements for collecting and analysing the data.

Asymptotically, unless the prior distribution assigns zero probability density to the true parameter values, the posterior distribution is proportional to the likelihood function,

$$\lim_{n \to \infty} g(\boldsymbol{\theta} \mid \mathcal{D}_n) \propto \lim_{n \to \infty} \mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_n). \tag{30}$$

From this asymptotic point of view, the analysis is robust against the choice of prior distribution and it is feasible to simplify matters considerably by assuming independent uniform priors or Jeffreys' invariant priors for the model parameters. In particular, this avoids the need to specify hyperparameters for subjective marginal prior distributions. Whether such a large-sample approximation is suitable depends on the modelling context. For example, if we were to record coordinates that identify the location of a football every second during a game, we would rapidly accumulate large amounts of data. Conversely, the times when goals are scored would yield sparse data sets and asymptotic results would not apply.

5. Decision analysis

The first aspect of decision analysis that we consider is outcome prediction. Depending upon what sports concern us, typical outcomes $\mathbf{y}$ and data $\mathbf{x}$ might comprise numbers or measures of strokes, scores, runs, wickets, points, games, goals, penalties, times, placings, baskets and possession. The posterior predictive distribution with probability mass or density function

$$f(\mathbf{y} \mid \mathcal{D}_n) = \int_{\theta} f(\mathbf{y} \mid \boldsymbol{\theta})g(\boldsymbol{\theta} \mid \mathcal{D}_n)d\boldsymbol{\theta} \qquad (31)$$

assigns exact probabilities to the random vector $\mathbf{y}$ given the observed data $\mathcal{D}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The simpler approximation

$$f(\mathbf{y} \mid \mathcal{D}_n) \approx f\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_n\right) \qquad (32)$$

based on a point estimate $\hat{\boldsymbol{\theta}}_n$ of the parameter vector $\boldsymbol{\theta}$ calculated from the likelihood function $\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_n)$ can lead to incorrect outcome predictions (see Bernardo and Smith, 1994).

The second aspect of decision analysis that we consider is strategy selection. Depending upon what sports concern us, typical strategies might comprise measures of amount of effort, degree of risk, attack or defence, substituting players, declaring innings or reviewing decisions. If strategy s_i has utility $u(s_i, \mathbf{y})$ for $i = 1, 2, \ldots, r$, then the best strategy maximises the posterior expected utility

$$E\{u(s_i, \mathbf{y}) \mid \mathcal{D}_n\} = \int_{\mathbf{y}} u(s_i, \mathbf{y})f(\mathbf{y} \mid \mathcal{D}_n)d\mathbf{y}. \qquad (33)$$

As explained by O'Hagan (1994), the simpler approximation

$$E\{u(s_i, \mathbf{y}) \mid \mathcal{D}_n\} \approx u(s_i, \hat{\mathbf{y}}_n) \qquad (34)$$

based on a point prediction $\hat{\mathbf{y}}_n$ of the random vector $\mathbf{y}$ calculated from the estimated sampling distribution $f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_n)$ can also lead to incorrect strategy selections.

Percy (2009) considered strategy selection and outcome prediction for badminton, and we elaborate upon this application now. This is a classical example of a discrete-time stochastic process, for which we observe a discrete random vector after each rally. The outcome Y that we wish to predict is whether team 1 loses ($y = 0$) or wins ($y = 1$) a particular game. The data $\mathbf{x}_n$ are the scores (n_1, n_2) of teams 1 and 2, respectively, after $n = n_1 + n_2$ rallies and the parameter θ is the probability that team 1 wins any specific rally. This simple assumption of constant θ throughout the game, regardless of which team or player serves, yields readily to analytical solution. However, simulation can be used to fit the model if we wish to relax this assumption to allow for variable θ.

Current game rules allow teams to score points on all serves and each game ends when either team scores 21 or more points with a lead of at least 2 points, subject to a maximum of 30 points. In either case, the team with the larger score wins the game. The transition matrix has the sparse form

	(0, 1)	(1, 0)	(0, 2)	(1, 1)	(2, 0)	$\cdots$	(29, 30)	(30, 29)
(0, 0)	$1-\theta$	θ	0	0	0	$\cdots$	0	0
(0, 1)	0	0	$1-\theta$	θ	0	$\cdots$	0	0
(1, 0)	0	0	0	$1-\theta$	θ	$\cdots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
(29, 29)	0	0	0	0	0	$\cdots$	$1-\theta$	θ

according to the recurrence relation

$$P(Y = 1 \mid n_1, n_2, \theta) = \theta P(Y = 1 \mid n_1 + 1, n_2, \theta)$$
$$+ (1-\theta)P(Y = 1 \mid n_1, n_2 + 1, \theta).$$

Percy (2009) solved this recurrence relation to find explicit forms for $P(Y = 1 | n_1, n_2, \theta)$ when the game score is (n_1, n_2). Conditional upon a total of $n_1 + n_2$ rallies having been played, the sampling distribution is

$$n_1 \mid n_1 + n_2, \theta \sim Bi(n_1 + n_2, \theta).$$

Using Jeffreys' invariant prior by default, a corresponding prior–posterior analysis then gives

$$\theta \sim Be\left(\frac{1}{2}, \frac{1}{2}\right) \Rightarrow \theta \mid n_1, n_2 \sim Be\left(n_1 + \frac{1}{2}, n_2 + \frac{1}{2}\right).$$

Finally, in order to calculate the evolving probabilities that team 1 wins a game as it progresses, we evaluate

$$P(Y = 1 \mid n_1, n_2) = \int_{-\infty}^{\infty} P(Y = 1 \mid n_1, n_2, \theta)g(\theta \mid n_1, n_2)d\theta$$

using the distributions defined above and a suitable numerical quadrature algorithm. This information is useful to the players for determining strategies, and to bookmakers and gamblers for determining in-play odds.

A flexible extension to this analysis involves a state space model that allows θ to vary over time. In this case, the sampling distribution (measurement equation) takes the form

$$n_1 \mid n_1 + n_2, \theta_t \sim Bi(n_1 + n_2, \theta_t)$$

and the corresponding transition equation becomes

$$\phi_{t+1} = \beta\phi_t + \eta_{t+1},$$

where

$$\phi_t = \log \frac{\theta_t}{1 - \theta_t}$$

and

$$\eta_t \sim No(a, b).$$

This model offers the flexibility whereby the probability that team 1 wins a rally is not constrained to be constant throughout the game. Rather, it is allowed to evolve continuously to reflect relative surges of energy or periods of lethargy among the players. The parameter β is an unknown, positive constant close to 1 and is readily incorporated in a Bayesian framework.

6. Case study

In order to demonstrate the preceding theory with a simple example, we consider the annual boat race between Oxford and Cambridge Universities in the UK. In particular, we investigate the problem of predicting which team will win next year's race with no information other than the winners of previous races. We analyse the results of the 160 mens' races from 1829 to 2014 inclusive, as taken from the website of Boat Race Company Ltd (2014). Of these races, Oxford won 78, Cambridge won 81 and one was drawn. Since draws are rare in boat races, we ignore this observation for simplicity in our illustrative analyses. The time scale is discrete to represent successive races.

An initial model for this scenario is a Bernoulli process, whereby the outcomes X_n are coded as 0 (Oxford wins) or 1 (Cambridge wins) for $n \in \mathbb{N}$ with probability mass function

$$p(x_n \mid \theta) = \theta^{x_n}(1-\theta)^{1-x_n} \quad ; \qquad x_n \in \{0, 1\}.$$

The parameter θ represents the probability that Cambridge wins any specific race. The simplest approach to inference evaluates the maximum likelihood estimate for θ based on this Bernoulli process, giving $\hat{\theta} = (81/159) \approx 0.51$. However, our analysis is based on the Bayes scheme in order to demonstrate the possibility of future dynamic updating, in the sense of the preceding methodology.

As θ is bounded, an appropriate prior distribution, which also has the desirable property of being natural conjugate, is the beta form with probability density function

$$g(\theta) = \frac{1}{B(a, b)} \theta^{a-1}(1-\theta)^{b-1}; \qquad \theta \in (0, 1).$$

With no specific prior knowledge other than a reasonable assumption of symmetry, we set $a = b = 1$ corresponding to the default uniform prior. Continual updating is not required to predict the result of race 161, as the posterior density also has a beta form

$$g(\theta \mid \mathcal{D}_{160}) = \frac{1}{B(82, 79)} \theta^{81}(1-\theta)^{78}; \qquad \theta \in (0, 1),$$

where $\mathcal{D}_{160} = \{x_1, \ldots, x_{160}\}$. Consequently, we need only to evaluate the posterior predictive probability mass function

$$p(x_{161} \mid \mathcal{D}_{160}) = \int_{-\infty}^{\infty} p(x_{161} \mid \theta)g(\theta \mid \mathcal{D}_{160})d\theta$$

$$= \frac{1}{B(82, 79)} \int_0^1 \theta^{81+x_{161}}(1-\theta)^{79-x_{161}}d\theta$$

$$= \frac{B(82+x_{161}, 80-x_{161})}{B(82, 79)}$$

$$= \begin{cases} \frac{79}{161} \approx 0.49; & x_{161} = 0 \quad \text{(Oxford wins)} \\ \frac{82}{161} \approx 0.51; & x_{161} = 1 \quad \text{(Cambridge wins)} \end{cases}.$$

These exact Bernoulli probabilities differ negligibly from those of the simpler approximation because we observed a large amount of data and chose to use an objective prior. In events where few data are available for analysis or a subjective prior is used, these differences can be significant.

However, we might reasonably expect last year's winner to have a greater chance of winning this year than has last year's loser. Part of the explanation is due to retaining particularly good athletes and team support, while part is due to the increased levels of confidence and morale that often accompany recent success. In order to allow for this dependency, we first extend the Bernoulli process into a simple form of discrete-time Markov chain with two states, 0 and 1 as before, and transition matrix

$$\mathbf{P} = \begin{pmatrix} \theta_0 & 1-\theta_0 \\ 1-\theta_1 & \theta_1 \end{pmatrix},$$

where the unknown parameters $\theta_0 = P(X_{n+1} = 0 | X_n = 0)$ and $\theta_1 = P(X_{n+1} = 1 | X_n = 1)$ are likely to be slightly more than one half for $n \in \mathbb{N}$. Glancing at the data also supports this suggestion of serial dependence, as the winning side won in the next race on 98 out of 158 (62%) of occasions. Indeed, the likelihood function that reflects this observation is given by

$$\mathcal{L}(\theta_0, \theta_1; \mathcal{D}_{160}) \propto \theta_0^{47}(1-\theta_0)^{30}\theta_1^{51}(1-\theta_1)^{30}$$

and the maximum likelihood estimates are $\hat{\theta}_0 = (47/77) \approx 0.61$ and $\hat{\theta}_1 = (51/81) \approx 0.63$. Again, though, we adopt the Bayes scheme for improved accuracy and to illustrate the dynamic updating aspects.

Referring to the generic forms in Equation (25), prior distributions that reflect the author's subjective knowledge are $\theta_i \sim Be(11, 7)$ for $i = 0, 1$ with $\theta_0 \amalg \theta_1$. As the data set is large, the results are robust against different choices of prior. For small data sets, knowledge would be elicited from experts and a sensitivity analysis would be performed. The corresponding joint prior probability density function has the form
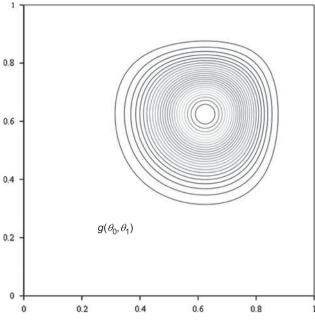
$$g(\theta_0, \theta_1) \propto \theta_0^{10}(1-\theta_0)^6\theta_1^{10}(1-\theta_1)^6; \qquad \theta_0, \theta_1 \in (0, 1)$$

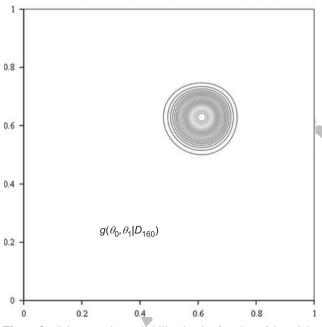and is illustrated by means of the contour plot in Figure 1.

By combining the likelihood and prior using Relation (26), the joint posterior then becomes

$$g(\theta_0, \theta_1 \mid \mathcal{D}_{160}) \propto \theta_0^{57}(1-\theta_0)^{36}\theta_1^{61}(1-\theta_1)^{36};$$

$$\theta_0, \theta_1 \in (0, 1)$$

from Relation (27). Hence, $\theta_0 \mid \mathcal{D}_{160} \sim Be(58, 37)$ and $\theta_1 \mid \mathcal{D}_{160} \sim Be(62, 37)$ with $\theta_0 \amalg \theta_1 \mid \mathcal{D}_{160}$. Figure 2 presents a contour plot of this joint posterior probability density function for comparison with the prior density in Figure 1. It clearly demonstrates the impact of the observed data upon our knowledge about the unknown transition probabilities, as the posterior density is more concentrated about its mode than is the prior density about its mode. That the modes are similarly located in

Figure 1   Joint prior probability density function of θ_0 and θ_1.

Figure 2   Joint posterior probability density function of θ_0 and θ_1.

both plots indicates that our original hunch about the carry-over effect for winning races was fairly accurate.

Noting that Oxford won the latest race, corresponding to $x_{160} = 0$, we have

$$p(x_{161} \mid x_{160}, \theta_0, \theta_1) = \theta_0^{1-x_{161}}(1-\theta_0)^{x_{161}}; \qquad x_{161} = 0, 1.$$

This enables us to predict the outcome of race 161 using Equation (31) to give

$$p(x_{161} \mid \mathcal{D}_{160}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_{161} \mid x_{160}, \theta_0, \theta_1)$$

$$g(\theta_0, \theta_1 \mid \mathcal{D}_{160})d\theta_0 d\theta_1$$

$$= \int_0^1 \frac{1}{B(58, 37)} \theta_0^{58-x_{161}}(1-\theta_0)^{36+x_{161}}$$

$$d\theta_0 \int_0^1 \frac{1}{B(62, 37)} \theta_1^{61}(1-\theta_1)^{36} d\theta_1$$

$$= \frac{B(59-x_{161}, 37+x_{161})}{B(58, 37)}$$

$$= \begin{cases} \frac{58}{95} \approx 0.61; & x_{161} = 0 \quad \text{(Oxford wins)} \\[2mm] \frac{37}{95} \approx 0.39; & x_{161} = 1 \quad \text{(Cambridge wins)} \end{cases}$$

Whereas the Bernoulli process marginally predicts that Cambridge will win the next race because it won the majority of previous races, this Markov chain predicts that Oxford will win because it depends primarily upon which team won the most recent race.

Before completing this case study, we make some further observations. First, the dynamic updating method of Relation (29) can be used in future years to avoid re-fitting the model from scratch each time. Of course, one race per year allows plenty of time in which to perform the calculations by either method. However, recent technological advances generate increasing numbers of situations that require similar dynamic decisions to be made instantaneously. Using dynamic updating when the outcome of next year's race is known, a prediction for the following year is obtained by evaluating

$$p(x_{162} \mid \mathcal{D}_{161}) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_{162} \mid x_{161}, \theta_0, \theta_1)\mathcal{L}(\theta_0, \theta_1; x_{160}, x_{161})$$

$$g(\theta_0, \theta_1 \mid \mathcal{D}_{160})d\theta_0 d\theta_1.$$

If $x_{161} = 0$, then

$$p(x_{162} \mid \mathcal{D}_{161}) \propto \int_0^1 \theta_0^{59-x_{162}}(1-\theta_0)^{36+x_{162}} d\theta_0$$

$$= \begin{cases} B(60, 37); & x_{162} = 0 \\[2mm] B(59, 38); & x_{162} = 1 \end{cases}$$

in which case $P(X_{162} = 0 \mid \mathcal{D}_{161}) = 0.61$ and $P(X_{162} = 1 \mid \mathcal{D}_{161}) = 0.39$ (Oxford wins/loses the subsequent race if it wins

the next race). Similarly, if $x_{161} = 1$, then

$$p(x_{162} \mid \mathcal{D}_{161}) \propto \int_0^1 \theta_1^{61+x_{162}}(1-\theta_1)^{37-x_{162}} d\theta_1$$

$$= \begin{cases} B(62, 38); & x_{162} = 0 \\ B(63, 37); & x_{162} = 1 \end{cases}$$

in which case $P(X_{162} = 0 \mid \mathcal{D}_{161}) = 0.37$ and $P(X_{162} = 1 \mid \mathcal{D}_{161}) = 0.63$ (Cambridge loses/wins the subsequent race if it wins the next race). Second, we note that one could also analyse the application in this case study by means of a time-series model or a state space formulation. The former would modify the Bernoulli process described above by introducing differencing or autoregressive terms to relate X_n to X_{n-1}. The latter would relate θ_n to θ_{n-1} more loosely than in the Markov chain considered here and in a similar manner as suggested for the badminton analysis of the preceding section. We plan to develop these ideas in detail for future presentation, in order to compare the predictive accuracies of the various modelling approaches.

7. Conclusions

The prevalence of technology appears to be ever increasing in sport, as a result of which many strategic decisions and outcome predictions are required in real time, during the course of play and often with little time for deliberation and calculation. We note that the literature has concentrated on modelling 'between-game' data and paid scant attention to 'within-game' data that arise when attempting to resolve this kind of problem. This paper reviews and categorises some of the most common modelling approaches in this context, based on discrete and continuous-time stochastic processes.

For discrete-time situations with discrete outcomes, some useful stochastic processes identified are Bernoulli processes, discrete-time Markov chains and hidden Markov models. With continuous outcomes, suitable stochastic processes include Gaussian random walks, times-series models and state space models. For continuous-time situations with discrete outcomes, some useful stochastic processes identified are continuous-time Markov chains, renewal processes and non-homogeneous Poisson processes. With continuous outcomes, suitable stochastic processes include Wiener processes, gamma processes and stochastic differential equations. In particular, we find that discrete-time Markov chains and state space models are the most common approaches because of their simple formulations and ease of implementation. However, we note that continuous-time Markov chains and various point process offer some advantages in terms of physical justification and modelling flexibility.

We also consider dynamic learning and decision analysis for these stochastic processes. First, we propose simple procedures for specifying objective and subjective prior distributions for unknown model parameters. Then we demonstrate how to calculate the posterior distribution using sequential Bayesian updating for efficiency. We claim that this provides a natural analytical framework that is easy to implement and leads to accurate decisions. We discuss two types of decision in this paper: outcome prediction and strategy selection. The former involves updating the posterior predictive distribution of the outcome measure dynamically, while the latter involves selecting the strategy that maximises posterior expected utility interactively during the course of play.

Finally, we present an interesting case study that illustrates sequential Bayesian updating for a Bernoulli process and a discrete-time Markov chain, with outcome prediction as the ultimate aim. This is a preliminary analysis to demonstrate the preceding concepts in practice and is based on simple assumptions. Nevertheless, this application shows that the calculations are fairly easy to perform and shows how effective this approach can be. Suggestions are made for improving the analysis by extending the model, should this or a similar application be of particular interest.

In the author's opinion, more research is needed to investigate time-varying vector outcomes in sport. Specifically, are there any benefits in adopting multivariate response distributions, taking account of robustness and increased parameterisation? Also, how might we choose among several possible forms of copula and how might we modify them dynamically to model dependence in random processes? Such modifications might incorporate knowledge relating to a competing risks setting such as that in cricket where a batsman can be out by good bowling, bad batting or random misfortune. Clearly, the study of stochastic processes in sport is itself dynamic and we must update our knowledge continually using the latest published research.

References

Barnett T and Clarke SR (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics* **16**(2): 113–120.

Bernardo JM and Smith AFM (1994). *Bayesian Theory*. Wiley: Chichester.

Boat Race Company Ltd (2014). Men's boat race results, http://theboatrace.org/men/results, accessed 24 September 2014.

Box G and Jenkins G (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day: San Francisco.

Brillinger DR (2007). A potential function approach to the flow of play in soccer. *Journal of Quantitative Analysis in Sports* online publication, 18 January, doi:10.2202/1559-0410.1048.

Cattelan M, Varin C and Firth D (2013). Dynamic Bradley-Terry modelling of sports tournaments. *Applied Statistics* **62**(1): 135–150.

Clarke SR (1988). Dynamic programming in one-day cricket—Optimal scoring rates. *Journal of the Operational Research Society* **39**(4): 331–337.

Clarke SR and Norman JM (1998). When to rush a 'behind' in Australian rules football: A dynamic programming approach. *Journal of the Operational Research Society* **49**(5): 530–536.

Congdon P (2003). *Applied Bayesian Modelling*. Wiley: Chichester.

Crowder M, Dixon M, Ledford A and Robinson M (2002). Dynamic modelling and prediction of English football league matches for betting. *The Statistician* **51**(2): 157–168.

Dalang RC, Dumas F, Sardy S, Morgenthaler S and Vila J (2014). Stochastic optimization of sailing trajectories in an upwind regatta. *Journal of the Operational Research Society* advance online publication, 7 May, doi:10.1057/jors.2014.40.

Dixon MJ and Coles SG (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* **46**(2): 265–280.

Dixon MJ and Robinson ME (1998). A birth process model for association football matches. *The Statistician* **47**(3): 523–538.

Duckworth FC and Lewis AJ (1998). A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society* **49**(3): 220–227.

Glickman ME and Stern HS (1998). A state-space model for national football league scores. *Journal of the American Statistical Association* **93**(441): 25–35.

Hirotsu N and Wright M (2002). Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society* **53**(1): 88–96.

Kemeny JG and Snell JL (1960). *Finite Markov Chains*. Springer-Verlag: New York.

Knorr-Held L (2000). Dynamic ratings of sports teams. *The Statistician* **49**(2): 261–276.

Knottenbelt WJ, Spanias D and Madurska AM (2012). A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Computers & Mathematics with Applications* **64**(12): 3820–3827.

Kostuk KJ and Willoughby KA (1999). 'Rocks' the 'house'. *OR/MS Today* **26**(6): 36–39.

Maher MJ (1982). Modelling association football scores. *Statistica Neerlandica* **36**(3): 109–118.

Maher MJ (2013). Predicting the outcome of the Ryder Cup. *IMA Journal of Management Mathematics* **24**(3): 301–309.

McGarry T and Franks IM (1994). A stochastic approach to predicting competition squash match–play. *Journal of Sports Sciences* **12**(6): 573–584.

McHale I and Scarf PA (2011). Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling* **11**(3): 219–236.

Nelsen RB (1999). *An Introduction to Copulas*. Springer-Verlag: New York.

O'Hagan A (1994). *Kendall's Advanced Theory of Statistics: Volume 2B Bayesian Inference*. Arnold: London.

Owen A (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics* **22**(2): 99–113.

Percy DF (2009). A mathematical analysis of badminton scoring systems. *Journal of the Operational Research Society* **60**(1): 63–71.

Percy DF (2011a). Interactive shrinkage methods for class handicapping. *IMA Journal of Management Mathematics* **22**(2): 139–156.

Percy DF (2011b). Prior elicitation: A compromise between idealism and pragmatism. *Mathematics Today* **47**(3): 142–147.

Percy DF, Kearney JR and Kobbacy KAH (2010). Hybrid intensity models for repairable systems. *IMA Journal of Management Mathematics* **21**(4): 395–406.

Pfeiffer M, Zhang H and Hohmann A (2010). A Markov chain model of elite table tennis competition. *International Journal of Sports Science and Coaching* **5**(2): 205–222.

Riddle LH (1988). Probability models for tennis scoring systems. *Applied Statistics* **37**(1): 63–75.

Sadovskiĭ LE and Sadovskiĭ AL (1993). *Mathematics and Sports*. Translated by Makar-Limanov S and edited by Ivanov SAmerican Mathematical Society: Providence.

Spanias D and Knottenbelt WJ (2013). Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics* **24**(3): 311–320.

Smith DK (2007). Dynamic programming and board games: A survey. *European Journal of Operational Research* **176**(3): 1299–1318.

Stefani RT (2009). Predicting score difference versus score total in rugby and soccer. *IMA Journal of Management Mathematics* **20**(2): 147–158.

Stern SE (2009). An adjusted Duckworth-Lewis target in shortened limited overs cricket matches. *Journal of the Operational Research Society* **60**(2): 236–251.

Urban TL (2012). The existence of serial correlation and its effect on the format of a championship playoff series. *Journal of the Operational Research Society* **63**(7): 883–889.

Volf P (2009). A random point process model for the score in sport matches. *IMA Journal of Management Mathematics* **20**(2): 121–131.

West M and Harrison J (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag: New York.

Wright M (1988). Probabilities and decision rules for the game of squash rackets. *Journal of the Operational Research Society* **39**(1): 91–99.