MODELLING OUTCOMES IN VOLLEYBALL

Barnett, Tristan [1], Brown, Alan [2] and Jackson, Karl [1]

[1] Faculty of Life and Social Sciences, Swinburne University of Technology, Australia

[2] Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Australia

Paper Submitted for Review ??/??/08 (editors will insert the dates)

Paper re-submitted for amendments ??/??/08 (include if re-submitted)

Abstract. A Markov Chain model is applied to volleyball to calculate win probabilities and mean lengths with the associated variances, conditional on both the scoreboard and the server. A feature of this model is that it predicts outcomes conditional on both the scoreboard and the serving team. The inclusion of the serving team in the event space is an essential requirement of this model, and arises from the rule in volleyball that the winner of each point in a set must serve on the following point. The average probability of a team winning a point on serve is less than 0.5, and so rotation of serve is commonplace. The key to the analysis of an evenly contested set is the observation that, from the situation where the scores are level (after at least 46 points have been played), the team that wins the set must eventually win two successive points. If the two points are shared then the score is level once again, although a rotation of server has occurred. This scoring structure, combined with the method of rotating the serve, distinguishes volleyball from other racket sports such as tennis, squash, badminton and table tennis. Results from the model indicate that it is advantageous to be the receiver on the opening point of a set and the team that wins the toss at the start of the fifth set (if the set score reaches 2-all), has an advantage for the remainder of the match. However, due to the rotation of serve after each set, there is no advantage for either side in being server or receiver at the start of the match.

Keywords: volleyball, Markov Chain model, scoring systems

INTRODUCTION

Markov Chain models are widely used in modelling sporting outcomes. Kemeny and Snell (1960) recognized that tennis could be modelled by Markov Chains. Tennis has four levels (point, game, set, match) and the time to play the match is not fixed, but rather depends on a player winning 2 or 3 sets. Other racket sports contain a similar structure to tennis and Markov Chain models could be developed to model these sports. For example, Clarke & Norman (1979) use Markov Chain models to compare North American and International squash scoring systems.

A Markov chain model can be applied to volleyball in a similar approach to racket sports. In volleyball, scoring consists of three levels: point, sets and match. A coin is tossed to determine the first serve of a match. At the start of each set, the team that was receiving first in the previous set becomes the server for the next set. If the set score reaches 2-all, then the toss of a coin decides the server for the start of the final set. Each team can win a point while either serving or receiving. The first team to win three sets wins the match. Each set is played as a 25-point set with the exception that after the set score reaches 2-all, the final set is played as a 15-point set. If the score reaches 24-all in a 25-point set, then play continues indefinitely until one team has obtained a two point lead. Similarly, if the score reaches 14-all in a 15-point set then the play continues indefinitely until one team has obtained a two point lead.

In this paper, a Markov Chain model is applied to volleyball to predict outcomes conditional on both the scoreboard and the server. The inclusion of the server in the event space is an essential feature of the model for volleyball because of the rule on serving in this sport, and distinguishes it from models for other racket sports such as tennis, squash, badminton and table tennis.

PROBABILITY OF WINNING

Point

Volleyball has the added complication of having 6 players that make up a team, rather than just the one player each side as occurs in racket sports. To simplify the analysis we will assume throughout that the probabilities of winning a point by each player in a team on their respective serves are identical and constant, irrespective of the score. Therefore, the model consists of two parameters, the probabilities of team A and team B winning a point on their respective serves. The probabilities of winning a point on serve are represented as follows:

	A winning point	B winning point
A serving	p_A	$q_A = 1 - p_A$
B serving	$q_B = 1 - p_B$	p_B

Set

The conditional probabilities of a team winning a set from point score (a, b) in a 25-point set are represented as follows:

	A winning set	B winning set
A serving next	P(A | A,a,b)	P(B | A,a,b)=1- P(A | A,a,b)
B serving next	P(A | B,a,b)	P(B | B,a,b)=1- P(A | B,a,b)

The conditional probabilities of a team winning a set from point score (a, b) in a 15-point set are represented as follows:

	A winning set	B winning set
A serving next	P*(A | A,a,b)	P*(B | A,a,b)=1- P*(A | A,a,b)
B serving next	P*(A | B,a,b)	P*(B | B,a,b)=1- P*(A | B,a,b)

We will adopt a similar notation for all other occasions where we need to distinguish between a 25-point set and a 15-pont set without further comment.

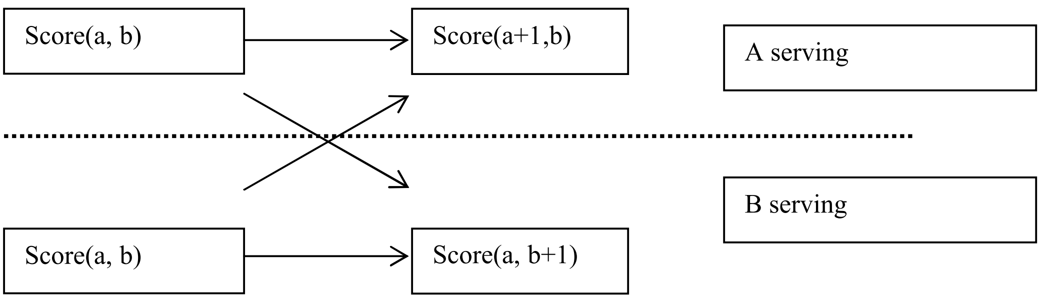

Figure 1. One step transitions between states of play.

Figure 1 illustrates the one step transitions between the various states of play. We set up a Markov chain model for the case of team A winning a 25-point set using backwards recurrence formulas, as follows:

$$P(A \mid A,a,b) = p_A\, P(A \mid A,a+1,b) + q_A\, P(A \mid B,a,b+1)$$
$$P(A \mid B,a,b) = p_B\, P(A \mid B,a,b+1) + q_B\, P(A \mid A,a+1,b)$$

The boundary values are:

$P(A \mid A,a,b) = 1$ if $a = 25, 0 \le b \le 23$
$P(A \mid B,a,b) = 0$ if $b = 25, 0 \le a \le 23$
$P(A \mid A,24,24) = p_A^2 / D$
$P(A \mid B,24,24) = p_A\, q_B\, (1+ p_A p_B - q_A\, q_B) / D$

where $D = (1 - q_A\, q_B)^2 - p_A\, q_A\, p_B\, q_B$

A method for determining the final boundary values, $P(A|A,24,24)$ and $P(A|B,24,24)$, is explained below. It is however straight forward to check using the recurrence formulas and boundary conditions that

$P(A \mid A,23,23) = P(A \mid A,24,24)$, and
$P(A \mid B,23,23) = P(A \mid B,24,24)$.

Similar formulas can be developed for a 15-point set. In the case where A wins a 15-point set the backwards recurrence formulas are

$P^*(A \mid A,a,b) = p_A\, P^*(A \mid A,a,+1,b)+ q_A\, P^*(A \mid B,a,b+1)$
$P^*(A \mid B,a,b) = p_B\, P^*(A \mid B,a,b+1) + q_B\, P^*(A \mid A,a+1,b)$

The boundary values are:

$P^*(A \mid A,a,b) = 1$ if $a = 15, 0 \le b \le 13$
$P^*(A \mid B,a,b) = 0$ if $b = 15, 0 \le a \le 13$
$P^*(A \mid A,24,24) = p_A^2 / D$
$P^*(A \mid B,24,24) = p_A\, q_B\, (1+ p_A p_B - q_A\, q_B) / D$

The formulas to cover the cases where B wins the set are obvious.

Table 1 represents the probability of team A winning a 25-point and 15-point set for different values of p_A and p_B from the start of the set. The average probability of winning points on serve in men's volleyball is about 0.25. Therefore the values of p_A and p_B were chosen to reflect this value. The results indicate that the team receiving first has an advantage in winning the set. This is not an unexpected result, since the receiving team has the first opportunity at an attack.

	25-point set		15-point set	
p_A, p_B	A serving	B serving	A serving	B serving
0.30, 0.30	0.48	0.52	0.47	0.53
0.30, 0.29	0.51	0.56	0.49	0.56
0.30, 0.25	0.63	0.69	0.59	0.66
0.25, 0.25	0.47	0.53	0.46	0.54
0.25, 0.24	0.50	0.57	0.48	0.57
0.20, 0.20	0.46	0.54	0.45	0.55

Table 1: The probability of team A winning a 25-point and 15-point set for different values of p_A and p_B from the start of the set.

Match

The probabilities of a team winning a 25-point set from its beginning are represented as follows:

	A winning set	B winning set
A serving first point of set	g_{AA} = P(A | A,0,0)	g_{AB} = P(B | A,0,0)=1- g_{AA}
B serving first point of set	g_{BA} = P (A | B,0,0)	g_{BB} = P(B | B,0,0)=1- g_{BA}

The conditional probabilities of a team winning a match from set score (c,d) are represented as follows:

	A winning match	B winning match
A to serve first point of set	G(A | A,c,d)	G(B | A,c,d)= 1 - G(A | A,c,d)
B to serve first point of set	G(A | B,c,d)	G(B | B,c,d)= 1 – G(A | B,c,d)
Toss to serve first point of set	G(A | *,c,d)	G(B | *,c,d)= 1 – G(A | *,c,d)

The boundary values for team A winning a match from set score (c, d) are given by

$G(A \mid A,c,d) = 1$ if $c = 3, 0 \le d \le 2$

$G(A \mid B,c,d) = 0$ if $d = 3, 0 \le c \le 2$

A toss for serve is required at the start of the match, and the serve rotates at the start of each subsequent set unless the set score reaches 2-all, when another toss for serve is required. Thus when (c, d) = (0, 0) or (c, d) = (2, 2) the formula for the toss is

$$G(A \mid *,c,d) = 0.5 * [\, G(A \mid A,c,d)+ G(A \mid B,c,d)\,]$$

The recurrence formulas after the first toss, and before the fifth set are

$$G(A \mid A,c,d) = g_{AA}\, G(A \mid B,c+1,d)+ g_{AB}\, G(A \mid B,c,d+1)$$

$$G(A \mid B,c,d) = g_{BB}\, G(A \mid A,c,d+1)+ g_{BA}\, G(A \mid A,c+1,d)$$

The recurrence formulas after the toss for the 15-point fifth set are

$$G(A \mid A,c,d) = g^*_{AA}\, G(A \mid B,c+1,d)+ g^*_{AB}\, G(A \mid B,c,d+1)$$

$$G(A \mid B,c,d) = g^*_{BB}\, G(A \mid A,c,d+1)+ g^*_{BA}\, G(A \mid A,c+1,d)$$

However when (c,d) = (2, 1) or (c,d) = (1, 2) one of the possible outcomes at the end of the set is a level score of 2 sets all, and recurrence formulas require modifications to allow for the toss.

$$G(A \mid A,2,1) = g_{AA}\, G(A \mid B,3,1)+ g_{AB}\, G(A \mid *,2,2)$$

$$G(A \mid B,2,1) = g_{BB}\, G(A \mid *,2,2)+ g_{BA}\, G(A \mid A,3,1)$$

$$G(A \mid A,1,2) = g_{AA}\, G(A \mid *,2,2)+ g_{AB}\, G(B \mid B,1,3)$$

$$G(A \mid B,1,2) = g_{BB}\, G(A \mid A,1,3)+ g_{BA}\, G(A \mid *,2,2)$$

When the boundary conditions are applied we obtain the simplification:

$$G(A \mid *,2,2) = 0.5 * [g^*_{AA} + g^*_{BA}]$$

The formulas for the probabilities of team B winning a match are obvious. Table 2 represents the probability of team A winning a match for different values of p_A and p_B from the start of the match. The results indicate that there is no advantage in serving or receiving at the start of the match. At the start of the fifth set, another coin toss is used to determine the serving team, and as given in Table 1, it is an advantage for the remainder of the match to be receiving first in this final set.

	match	
p_A, p_B	A serving	B serving
0.30, 0.30	0.50	0.50
0.30, 0.29	0.56	0.56
0.30, 0.25	0.77	0.77
0.25, 0.25	0.50	0.50
0.25, 0.24	0.56	0.56
0.20, 0.20	0.50	0.50

Table 2: The probability of team A winning a match for different values of p_A and p_B from the start of the match.

Finding the boundary conditions at the end of a set

The key to determining boundary values such as P(A | A,24,24) and P(A | B,24,24) , is to observe that, from the situation the scores are level when at least 46 points have been played, the team that wins the set must eventually win two successive points. If the two successive points are shared then the score is level once again, although a rotation of server may have taken place. Figure 2 illustrates the one step transitions between the various states of play after the scores are level.

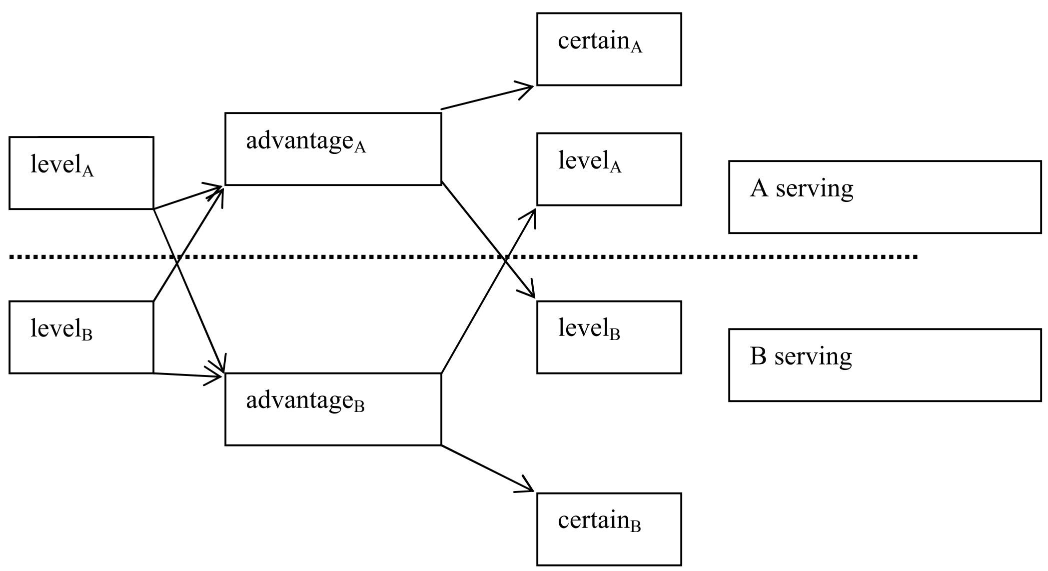

Figure 2. One step transitions between states of play, in a set after scores are level and when at least 46 points already played.

From the independence of the outcome of successive points we can calculate the two step transitions, and use this to eliminate the advantage states from the diagram, as shown in Figure 3.

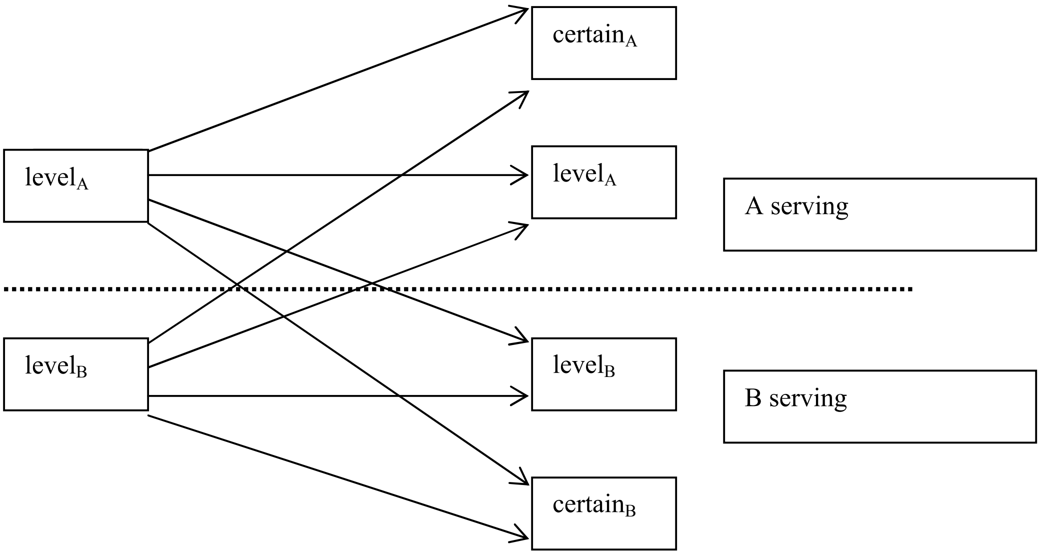

Figure 3. Two step transitions between states of play, in a set after scores are level and when at least 46 points already played.

We consider the case of A winning the set after reaching various states. To do this we must set the appropriate boundary conditions:

$$c_A = P(A \mid A,25,23) = P(A \mid B,25,23) = 1,$$

and

$$c_B = P(A \mid A,23,25) = P(A \mid B,23,25) = 0.$$

To simplify the notation let

$w_A = P(A \mid A,24,24)$, and $w_B = P(A \mid B,24,24)$.

Since

$$w_A = P(A \mid A,23,23) = P(A \mid A,24,24) \quad \{ = P(A \mid A,25,25) = \ldots\ldots \}$$

and

$$w_B = P(A \mid B,23,23) = P(A \mid B,24,24) \quad \{ = P(A \mid B,25,25) = \ldots\ldots \}$$

we can use the backwards recurrence relations already given to obtain, after two steps

$$w_A = p_A^2\, c_A + q_A q_B\, w_A + p_A q_A\, w_B + q_A p_B\, c_B$$
$$w_B = q_B p_A\, c_A + p_B q_B\, w_A + q_A q_B\, w_B + p_B^2\, c_B$$

When the boundary conditions are taken into account these equations simplify to

$$w_A (1 - q_A q_B) - p_A q_A\, w_B = p_A^2$$
$$w_B (1 - q_A q_B) - p_B q_B\, w_A = p_A q_B$$

Solving this pair of simultaneous equations leads to

$$w_A = p_A^2 / [(1 - q_A q_B)^2 - p_A q_A p_B q_B]$$
$$w_B = p_A q_B (1 + p_A p_B - q_A q_B) / [(1 - q_A q_B)^2 - p_A q_A p_B q_B]$$

The same argument can be used to develop the boundary conditions for the case where team A wins a 15-point set, leading to identical results. Results for the cases where team B wins a set can be obtained by symmetry, using similar arguments.

NUMBER OF POINTS PLAYED IN A SET

Mean number of points in a set

The mean number of points remaining in a 25-point set from point score (a, b) are represented as follows:

A serving next	$M(A\|a,b)$
B serving next	$M(B\|a,b)$

The backwards recurrence formulas are as follows:

$$M(A|a,b) = 1 + p_A M(A|a+1,b) + q_A M(B|a,b+1)$$
$$M(B|a,b) = 1 + p_B M(B|a,b+1) + q_B M(A|a+1,b)$$

The boundary values are:

$$M(A| a,b) = 0 \quad \text{if } a = 25,\ 0 \leq b \leq 23$$
$$M(B|a,b) = 0 \quad \text{if } b = 25,\ 0 \leq a \leq 23$$
$$M(A|24,24) = 2[1 + p_A q_A - q_A q_B]/D$$
$$M(B|24,24) = 2[1 + p_B q_B - q_A q_B]/D$$

where $D = (1 - q_A q_B)^2 - p_A p_B q_A q_B$

Variance of the number of points in a set

The variance of the number of points remaining in a 25-point set from point score (a, b) are represented as follows:

A serving next	$V(A\|a,b)$
B serving next	$V(B\|a,b)$

The backwards recurrence formulas are as follows:

$$V(A|a,b) = p_A V(A|a+1,b) + q_A V(B|a,b+1) + p_A q_A [(M(A|a+1,b) + M(B|a,b+1)]^2$$
$$V(B|a,b) = p_B V(B|a,b+1) + q_B V(A|a+1,b) + p_B q_B [M(B|a,b+1) + M(A|a+1,b)]^2$$

The boundary values are:

$$V(A|a,b) = 0 \quad \text{if } a = 25,\ 0 \leq b \leq 23$$
$$V(B|a,b) = 0 \quad \text{if } b = 25,\ 0 \leq a \leq 23$$
$$V(A|24,24)=4q_A[p_A+q_B+3p_Ap_Bq_B-2q_Aq_B^2-p_A^2q_A-p_Ap_Bq_Aq_B^2 - p_Aq_A^2q_B^2 + p_A^2p_Bq_Aq_B + q_A^2q_B^3]/D^2$$
$$V(B|24,24)=4q_B[p_B+q_A+3p_Ap_Bq_A-2q_A^2q_B-p_B^2q_B-p_Ap_Bq_A^2q_B - p_Bq_A^2q_B^2 + p_Ap_B^2q_Aq_B + q_A^3q_B^2]/D^2$$

Similar recurrence formulas can be devised for the mean and variance of the number of points remaining in a 15-point set and the mean and variance of the number of sets remaining in a match. Table 3 gives numerical results of the number of points in a 25-point set for different values of p_A and p_B.

	Mean points in 25-point set		Standard deviation of the number of points in a 25-point set	
p_A, p_B	A serving	B serving	A serving	B serving
0.30, 0.30	47.0	47.0	4.7	4.7
0.30, 0.29	47.1	47.1	4.8	4.8
0.30, 0.25	47.2	47.0	5.0	5.0
0.25, 0.25	47.8	47.8	5.4	5.4
0.25, 0.24	47.9	47.8	5.5	5.5
0.20, 0.20	48.9	48.9	6.7	6.7

Table 3: The mean and standard deviation of the number of points in a 25-point set for different values of p_A and p_B.

CONCLUSIONS

This paper has demonstrated that the use of Markov chains can be used to model outcomes in volleyball conditional on both the scoreboard and the server. Results from the model indicate that it is advantageous to be the receiver on the opening point of a set and the team that wins the toss at the start of the fifth set (if the set score reaches 2-all), has an advantage for the remainder of the match. However, due to the rotation of serve after each set, there is no advantage for either side in being server or receiver at the start of the match. Similar models could also be applied to beach volleyball, where the rotation of serve in beach volleyball is the same as standard volleyball.

References

Clarke, S.R. and Norman, J.M. (1979) Comparison of North American and international squash scoring systems-analytical results. *Research Quarterly* **50(4)**: 723–728.

Kemeny, J.G. and Snell, J.L. (1960) *Finite Markov chains*. Princeton, New Jersey, USA.