APPLICATIONES MATHEMATICAE
33,1 (2006), pp. 51–59

I. Kopocińska and B. Kopociński (Wrocław)

CUMULATIVE PROCESSES IN BASKETBALL GAMES

Abstract. We assume that the current score of a basketball game can be modeled by a bivariate cumulative process based on some marked renewal process. The basic element of a game is a cycle, which is concluded whenever a team scores. This paper deals with the joint probability distribution function of this cumulative process, the process describing the host's advantage and its expected value. The practical usefulness of the model is demonstrated by analyzing the effect of small modifications of the model parameters on the outcome of a game. The 2001 Lithuania–Latvia game is used as an example.

Introduction. The basic element of a basketball game defined in this paper is a cycle of a game. Dembiński and Kopociński (2003) define a cycle as a period in which just one team has possession of the ball. They define appropriate stochastic processes and the stream of game events according to this definition. In this way, they observe a correlation between the characteristics of cycles. In this paper an alternative definition of a cycle based on the times at which teams score points is introduced. In order to test this model, we used many video-recorded games from the Suproliga, Saporta Cup and the Polish basketball league. This paper uses the 2001 Lithuania–Latvia game as an illustration. The final score was 77:94.

Model. We divide the course of a basketball game into cycles separated by the moments at which points are scored. A cycle is characterized by a pair of random variables (X, T), where X denotes the number of points scored in the cycle and T denotes the duration of the cycle. In this paper, we characterize the sequence of cycles in a basketball game, estimate the parameters of the sequence and analyze the cumulative processes of points scored. The practical usefulness of the model is demonstrated by numerically

2000 *Mathematics Subject Classification*: Primary 60G35; Secondary 60K20.

Key words and phrases: applied probability, Markov processes, sports, cumulative processes.

analyzing the effect of small modifications in the model parameters on the value of the cumulative process.

Let one team (the home team) be indexed by 0 and the other team (the away team) be indexed by 1. Let a binary sequence $\{U_n, n \geq 0\}$ describe the state of a cycle: $U_n = 0$ if the home team scores in the nth cycle and $U_n = 1$ if the visiting team scores in the nth cycle. Let U_0 be the initial state of the sequence. Let X_n denote the number of points scored in a cycle. Set $W_n = X_n$ for $U_n = 0$ and $W_n = -X_n$ for $U_n = 1$ (i.e. W_n is the advantage gained by the home team in the nth cycle). Define $W_n^{(u)}$ to be the random variable W_n under the condition $U_n = u$. We have $X_n = (1 - U_n)W_n^{(0)} - U_n W_n^{(1)}$. Let T_n denote the integer-valued duration of a cycle. The units of time are taken to be seconds. The duration of the game is 2400 units (40 minutes).

In modelling a basketball game, it may be supposed that $\{(X_n, T_n)\}$, $n \geq 1$, is a Markovian sequence. Unfortunately, the large number of possible states of the sequence reduces the chance of deriving a model which is useful in practice. Here we make the simplifying assumption that $\{U_n\}$ is a Markovian binary sequence and X_n and T_n depend only on U_n. Let $T_n^{(u)} = T_n \,|\, U_n = u$, $u = 0, 1$.

Let $(p_0^{(n)}, p_1^{(n)})$ denote the probability distribution function of $\{U_n\}$ and let $(p_{uk},\ u, k = 0, 1)$ denote the elements of the transition probability matrix, assuming that it is homogeneous.

The stationary Markovian sequence $\{U_n\}$ has two parameters defining a one-step transition matrix, namely $p_{00} = P(U_{n+1} = 0 \,|\, U_n = 0)$, $p_{11} = P(U_{n+1} = 1 \,|\, U_n = 1)$. Thus, $p_{01} = 1 - p_{00}$, $p_{10} = 1 - p_{11}$. Using the empirically observed transition matrix presented in Table 1, we can estimate the stationary probabilities:

$$p_0 = \frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \qquad p_1 = 1 - p_0.$$

In our model of a basketball game we also assume that $W_n^{(u)} \stackrel{d}{=} W^{(u)}$, $n \geq 0$, as well as $T_n^{(u)} \stackrel{d}{=} T^{(u)}$, $n \geq 0$, are independent, identically distributed random variables and $W_n^{(u)}$, $T_n^{(u)}$, $u = 0, 1$, are mutually independent for each n; $\stackrel{d}{=}$ denotes equality in distribution.

The regulations of the game do not preclude a large number of points in a cycle, but in practice scores of more than three points per cycle are not observed. Allowing the possibility of different probability distribution functions for the number of points scored in a cycle and the duration of cycles depending on whether the home or visiting team scores, we introduce the notation

$$P(W^{(u)} = i) = w_i^{(u)}, \qquad i \geq 1,$$
$$P(T^{(u)} = i) = t_i^{(u)}, \qquad i \geq 1,\ u = 0, 1.$$

Problem of parameter estimation. Let $\mathcal{S} = \{0, 1, 00, 01, 10, 11, 000, 100, \ldots\}$ denote the set of binary sequences. Table 1 shows the empirically observed frequency n_s of selected sequences for the game examined. The corresponding transition probability matrix for the binary sequence is: $p_{00} = 0.324$, $p_{01} = 0.676$, $p_{10} = 0.632$, $p_{11} = 0.368$. Note that the winner scores in a larger number of cycles and transition from one state of the cycle to the other is more likely than remaining in the same state. Therefore, in a basketball game possession of the ball after a score gives an advantage similar to serving in tennis.

Table 1. Number of given subsequences $s \in S$ of states in the sequence $\{U_n\}$ and testing the hypothesis that the sequence is Markovian for the Lithuania–Latvia game

s	0	1		00	01	10	11	
n_s	37	39		11	23	24	14	
s	000	001	010	011	100	101	110	111
n_s	5	5	14	8	6	16	9	5
$\bar{n}_s$	3.4	7.2	14.0	8.2	7.2	15.0	8.2	4.8

We test the hypothesis that the sequence $\{U_n\}$ is Markovian using triples of consecutive cycles. The probabilities p_{ijk} for $i, j, k = 0, 1$ can be calculated from the formula

$$p_{ijk} = p_i p_{ij} p_{jk}.$$

The expected numbers $\overline{n}_s$ of triples of consecutive cycles are presented in Table 1. The *chi-square* statistic for the goodness of fit test is equal to $\chi^2 = 1.73$. Since this statistic has $df = 7$ degrees of freedom, we do not reject this hypothesis for the game observed.

Table 2. Distribution $(w_i^{(u)})$, $u = 0, 1$, of the number of points scored in a cycle for the Lithuania–Latvia game

Variable	N	p_0	$w_1^{(0)}$	$w_2^{(0)}$	$w_3^{(0)}$	$w_1^{(1)}$	$w_2^{(1)}$	$w_3^{(1)}$
Value	72	0.483	0.135	0.649	0.216	0.103	0.385	0.513

Table 2 shows the number N of cycles in the game observed, the stationary probability p_0 and the empirical distributions of the number of points scored in a cycle. The *chi-square* statistic for the goodness of fit test for the hypothesis that the probability distribution functions $(w_i^{(0)})$ and $(w_i^{(1)})$ are identical is equal to $\chi^2 = 7.28$. Since there are $df = 2$ degrees of freedom, this leads to rejection of this hypothesis for the game observed.

The video records of a game permit the observation of time exact to within one second. Table 3 shows the empirical conditional probability distribution functions $(t_j^{(0)})$ and $(t_j^{(1)})$ for the cycle duration grouped in 10-second

intervals, together with some modifications of these distributions. In this way we obtain just a few classes for the distribution $t_j^{(u)} = P(\lceil \frac{1}{10} T_n \rceil = i \,|\, U_n = u)$, $i \geq 1$, $u = 0, 1$, where $\lceil\ \rceil$ is the largest integer function.

The *chi-square* statistic of the goodness of fit test for the hypothesis that these distributions are identical is equal to $\chi^2 = 17.48$. Since there are $df = 5$ degrees of freedom, this leads to rejection of this hypothesis for the game observed.

Cumulative processes of points scored. Let $\{Z_k(t), t \geq 0\}$, $k = 0, 1$, denote the cumulative processes of points scored, where k is the team index, and let $\{Z_k^{(u)}(t)\}$, $u = 0, 1$, denote the conditional processes given $U_0 = u$. Because the basic sequence $\{(X_n, T_n)\}$ is a marked Markov process, the analysis of the cumulative processes is well known (cf. Çinlar (1975), Kopocińska (1990)).

Let events $A_{uk} = \{U_{n+1} = k \,|\, U_n = u\}$, $u, k = 0, 1$, denote the transition from state u to state k in $\{U_n\}$ assuming homogeneity. Henceforth, we denote the indicator of A by I_A. Note that

$$Z_k(t) = I_{U_0=0} Z_k^{(0)}(t) + I_{U_0=1} Z_k^{(1)}(t).$$

PROPOSITION 1. *The cumulative processes* $\{Z^{(u)}\}$, $u = 0, 1$, *of points scored satisfy the following recurrence relations*:

$$Z_0^{(u)}(0) = 0, \qquad Z_1^{(u)}(0) = 0,$$

$$(Z_0^{(u)}(t), Z_1^{(u)}(t)) \stackrel{d}{=} \begin{cases} (0,0) & \text{if } A_{u0} \cap \{T^{(0)} > t\} \cup A_{u1} \cap \{T^{(1)} > t\}, \\ (W^{(0)} + Z_0^{(0)}(t - T^{(0)}), Z_1^{(0)}(t - T^{(0)})) & \text{if } A_{u0} \cap \{T^{(0)} \leq t\}, \\ (Z_0^{(1)}(t - T^{(1)}), W^{(1)} + Z_1^{(1)}(t - T^{(1)})) & \text{if } A_{u1} \cap \{T^{(1)} \leq t\}, \end{cases}$$

where $W^{(0)}$, $Z_0^{(0)}$, $Z_1^{(0)}$, A_{n0} *as well as* $W^{(0)}$, $Z_0^{(0)}$, $Z_1^{(0)}$, A_{n1} *are mutually independent*, $u = 0, 1$, $t \geq 0$.

PROPOSITION 2. *The following recurrence formulas hold for the probability distribution functions of the cumulative processes*:

$$P(Z_0^{(u)}(0) = n_0,\, Z_1^{(u)}(t) = n_1) = I_{n_0+n_1=0},$$

$$\begin{aligned} P(Z_0^{(u)}(t) = n_0,\, Z_1^{(u)}(t) = n_1) &= I_{n_0+n_1=0}(p_{u0} P(T^{(0)} > t) + p_{u1} P(T^{(1)} > t)) \\ &\quad + I_{n_0>0} p_{u0} \sum_{i=1}^{n_0} P(W^{(0)} = i) \sum_{j=1}^{t} P(T^{(0)} = j) \\ &\quad \times P(Z_0^{(0)}(t-j) = n_0 - i,\, Z_1^{(0)}(t-j) = n_1) \end{aligned}$$

$$+ I_{n_1>0} p_{u1} \sum_{i=1}^{n_1} P(W^{(1)} = i) \sum_{j=1}^{t} P(T^{(1)} = j)$$

$$\times P(Z_0^{(1)}(t-j) = n_0,\ Z_1^{(1)}(t-j) = n_1 - i)),$$

where $u = 0, 1,\ n_0 \geq 0,\ n_1 \geq 0,\ t \geq 0.$

Let us denote the cumulative processes describing the point advantage of the host team in the game under the condition $U_0 = u$ by

$$Z^{(u)}(t) = Z_0^{(u)}(t) - Z_1^{(u)}(t), \qquad u = 0, 1,\ t \geq 0.$$

PROPOSITION 3. *The cumulative processes describing the point advantage of the host team satisfy the following recurrence relations*:

$$Z^{(u)}(0) = 0,$$

$$Z^{(u)}(t) \stackrel{d}{=} \begin{cases} 0 & \text{if } A_{u0} \cap \{T^{(0)} > t\} \cup A_{u1} \cap \{T^{(1)} > t\}, \\ \sum_{k=0}^{1} I_{A_{uk}} I_{T^{(k)} \leq t}((-1)^k W^{(k)} + Z^{(k)}(t - T^{(k)})) & \text{otherwise.} \end{cases}$$

PROPOSITION 4. *The following recurrence relations hold for the probability distribution functions of the processes describing the point advantage of the host team*:

$$P(Z^{(u)}(0) = n) = I_{n=0},$$

$$P(Z^{(u)}(t) = n) = I_{n=0}(p_{u0} P(T^{(0)} > t) + p_{u1} P(T^{(1)} > t))$$

$$+ p_{u0} \sum_{i=1}^{n} P(W^{(0)} = i) \sum_{j=1}^{t} P(T^{(0)} = j) P(Z^{(0)}(t-j) = n - i)$$

$$+ p_{u1} \sum_{i=1}^{\infty} P(W^{(1)} = i) \sum_{j=1}^{t} P(T^{(1)} = j) P(Z^{(1)}(t-j) = n + i),$$

where $u = 0, 1,\ n = 0, \pm 1, \ldots,\ t \geq 1.$

We introduce the following notation for the expected values of the cumulative processes of scored points and the host's advantage: $M_k^{(u)}(t) = \mathrm{E}(Z_k^{(u)}(t))$, $M^{(u)}(t) = \mathrm{E}(Z^{(u)}(t))$, $t \geq 0$, under the condition $U_0 = u$, $u = 0, 1$. Let $\overline{W}^{(k)} = \mathrm{E}(W^{(k)})$, $u, k = 0, 1$.

PROPOSITION 5. *The following recurrence formulas are satisfied*:

1) *for the expected values of the cumulative processes of points scored*;

$$M_k^{(u)}(0) = 0,$$

$$M_0^{(u)}(t) = \sum_{l=0}^{1} p_{ul} \Big(\overline{W}^{(l)} k P(T^{(l)} \leq t) + \sum_{j=1}^{t} P(T^{(l)} = j) M_k^{(l)}(t-j) \Big),$$

2) *for the expected value of the process describing the host's advantage*:

$$M^{(u)}(0) = 0,$$

$$M^{(u)}(t) = \sum_{l=0}^{1} p_{ul}\Big((-1)^k \overline{W}^{(l)} P(T^{(l)} \le t) + \sum_{j=1}^{t} P(T^{l)} = j) M^{(l)}(t-j)\Big),$$

where $u, k = 0, 1$, $t \ge 1$.

The asymptotic behaviour of the cumulative processes defined for these marked renewal processes is well known (see for example Kopocińska (2001)). The most important term of the asymptotical expansions for the expected values depends upon the expected values of the variables describing a cycle. Hence, estimators of these parameters should be unbiased. The distribution is asymptotically normal, with second moments appearing, and the initial condition for the sequence of cycles (defining who has initial possession of the ball) is of little importance.

As stated previously, in this analysis the time duration of a cycle was grouped into 10-second intervals. Specifically, now we consider $\hat{T} = \lceil \frac{1}{10} T_n \rceil$. Unfortunately, under such a transformation, the relation between the expected values $10\mathrm{E}(\hat{T}) \sim \mathrm{E}(T)$ may not be precise enough from the point of view of Markov renewal theory. This leads to the following numerical problem: find a suitable transformation $\hat{T} :\to \breve{T}$ such that $10\mathrm{E}(\breve{T}) = \mathrm{E}(T)$ and the probability distribution functions of $\hat{T}$ and $\breve{T}$ are as similar as possible.

This problem may be solved in the following way. Let us consider the distribution $P(T = j) = t_j$, $j \in \mathcal{T}$, where $\mathrm{E}\big(\frac{1}{10}T\big) = \mu_1$. We seek a random variable $\breve{T}$ with distribution $P(\breve{T} = j) = \breve{t}_j$, $j \in \breve{\mathcal{T}}$, such that

$$\sum_{j \in \breve{\mathcal{T}}} \breve{t}_j = 1, \quad \sum_{j \in \breve{\mathcal{T}}} j\breve{t}_j = \mu_1, \quad \sum_{j \in \breve{\mathcal{T}}} (t_j - \breve{t}_j)^2 = \min.$$

Table 3. Discrete distribution $(t_j^{(u)})$, $u = 0, 1$, of the duration of a cycle and its transformations for the Lithuania–Latvia game

	0	1	2	3	4	5	6	≥ 6
$t_j^{(0)}$	0.054	0.243	0.243	0.189	0.081	0.108	0.054	0.028
$\breve{t}_j^{(0)}$	0.000	0.199	0.222	0.186	0.101	0.151	0.121	0.000
$\tilde{t}_j^{(0)}$	0.000	0.222	0.238	0.188	0.095	0.137	0.100	0.000
$t_j^{(1)}$	0.026	0.359	0.205	0.179	0.128	0.051	0.000	0.052
$\breve{t}_j^{(1)}$	0.000	0.292	0.169	0.171	0.152	0.107	0.089	0.000
$\tilde{t}_j^{(0)}$	0.000	0.313	0.183	0.173	0.147	0.095	0.069	0.000

In the example presented we obtain the expected values $\mu_1^{(0)} = \mathrm{E}(T^{(0)}) = 3.1757$, $\mu_1^{(1)} = \mathrm{E}(T^{(0)}) = 2.9103$.

The numerical solution of this problem is not difficult. Table 3 presents the empirical distributions $(t_j^{(u)})$, the transformed distributions $(\check{t}_j^{(u)})$ with a smaller support and the same expected value, as well as the transformed distribution $(\tilde{t}_j^{(u)})$, which has a 10% smaller expected value.

In our calculations we assume that $P(U_0 = 0) = P(U_0 = 1) = 0.5$. The probability distribution function $(w_i^{(u)})$ is taken from Table 1 and $(t_j^{(u)})$ from Table 3 ($(\check{t}_j^{(u)})$ and $(\tilde{t}_j^{(u)})$ are also used as alternatives). We calculate the probability distribution function of the random variable $Z_k(t) = \frac{1}{2}(Z_k^{(0)}(t) + Z_k^{(1)}(t))$, $u, k = 0, 1$, $0 \le t \le 20$.

Applications. Note that, in practice, a basketball game may be split into any number of fragments defined by the coaches' decisions regarding the lineup of teams or tactics in the game. Therefore, inference concerning the result of the whole game based on the analysis presented may be a poor approximation of reality. Analysis of such fragments of the game may well be interesting and important.

The joint distribution of the cumulative processes $(Z_k^{(u)})$, $u = 0, 1$, may be exploited to calculate the marginal probability distribution functions of $z_k(i) = P(Z_k(t) = i)$, $i \ge 0$, $k = 0, 1$, the distribution of the host's advantage $z(i) = P(Z_0(t) - Z_1(t) = i)$, $i = 0, \pm 1, \ldots$, (see Table 4) and the probability of the host's winning. Also, we calculate the moments of the joint distribution, in particular the correlation between the numbers of points scored by both teams (in the example given $\mathrm{Corr}(Z_0(20), Z_1(20)) = 0.333$). The large variance of the distribution of the cumulative processes explains the large variability of the results of the four quarters in a game, which is observed in practice.

Propositions 1–5 enable a study of the influence of small changes in the parameters of the model on the result of a fragment of a game. Here we consider modification of the transition probabilities in the sequence $\{U_n\}$, $u = 0, 1$, the probability distribution functions of shot effectiveness $(w_k^{(u)})$ and the expected duration of a cycle. We change these parameters rather arbitrarily (see Table 5). We increase the expected duration of a cycle by 10%. Note that the parameters of the model, treated as random variables for each team in a basketball league, are correlated. The time elapsed in a game may affect the effectiveness of actions. Let us simplify the problem by assuming that an error in the estimation of one variable does not cause errors in the estimation of others.

Our numerical results are as follows: in the game examined the expected numbers of points scored by the teams in 200 (i.e. $t = 20$) seconds are $\mathrm{E}(Z_0) = 5.561$, $\mathrm{E}(Z_1) = 6.941$ and the variances are $\mathrm{Var}(Z_0) = 10.877$,

Table 4. Distribution of the number of points scored and the host advantage for the Lithuania–Latvia game

n	$z_0(n)$	$z_1(n)$	$z(-n)$	$z(n)$
0	0.004	0.003	0.072	0.072
1	0.008	0.005	0.065	0.075
2	0.041	0.019	0.056	0.075
3	0.051	0.037	0.046	0.070
4	0.113	0.048	0.036	0.064
5	0.115	0.087	0.027	0.055
6	0.135	0.098	0.019	0.046
7	0.123	0.096	0.013	0.037
8	0.096	0.113	0.008	0.028
9	0.072	0.094	0.005	0.021
10	0.047	0.074	0.003	0.015
11	0.029	0.066	0.002	0.010
12	0.016	0.044	0.001	0.007
13	0.009	0.030	0.000	0.004
14	0.004	0.022	0.000	0.003
15	0.002	0.013	0.000	0.002
16	0.001	0.008	0.000	0.001
17	0.000	0.005	0.000	0.001
18	0.000	0.002	0.000	0.000
19	0.000	0.001	0.000	0.000
20	0.000	0.001	0.000	0.000

Table 5. Modifications of the parameters in the model of the Lithuania–Latvia game and the effect on the game result

Parameter	Increase in parameter	Increase in $\mathrm{E}(Z_0)$	Increase in $\mathrm{E}(Z_1)$
p_{00}	-0.1	0.399	0.498
p_{00}	0.1	0.454	-0.569
p_{11}	-0.1	0.425	-0.525
p_{11}	0.1	0.489	0.587
$w_1^{(0)}$	0.05	-0.20	0.00
$w_3^{(0)}$	0.05	0.20	0.00
$w_1^{(1)}$	0.05	-0.21	0.00
$w_3^{(1)}$	0.05	0.00	0.21
$\mathrm{E}(T^{(0)})$	0.1588	0.156	0.152
$\mathrm{E}(T^{(1)})$	0.1455	0.120	0.180

$\mathrm{Var}(Z_1) = 16.761$. Table 5 shows how parameter changes in the model affect the number of points scored in the game. We see that some parameters have a greater effect than others. These numerical results may well indicate to a coach profitable goals for training and how to select players for a game.

References

E. Çinlar (1975), *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ.

J. Dembiński and B. Kopociński (2003), *Modelling of the course of a basketball match*, Human Movement 2, no. 8, 11–15 (Polish, English summary).

I. Kopocińska (1990), *Regenerative renewal processes*, Zastos. Mat. 20, 329–343.

I. Kopocińska (2001), *Estimation of cumulative processes*, Complex Systems 13, 177–183.

Mathematical Institute
Wrocław University
Pl. Grunwaldzki 2/4
50-384 Wrocław, Poland
E-mail: ibk@math.uni.wroc.pl

Received on 16.1.2006;
revised version on 6.3.2006 (1802)