Non-markov discrete random walks - selected properties

April 20, 2020

Abstract

This paper elaborates the newly introduced model of discrete random walks with memory implemented through a memory coefficient λ. It follows on the recent work of the authors and further examines the theoretical properties of the model.

1 Introduction

Modelling sporting events and reliability analysis are two seemingly very different areas of mathematics, yet they offer many similarities. The key factor in both domains is a certain probability measure determining the development of observed phenomenon. This probability measure is called "hazard rate" in reliability analysis and simply "probability of scoring" in modelling of sport events. It is altered either continuously with time according to some underlying probability function or suddenly as a reaction to an event. Breakage or repair of a given machine in reliability analysis or scoring (goal, point, basket etc.) in sports modelling. The most prominent tool in reliability analysis is the Cox's model ZDROJ expressing the changes of a hazard rate with a base hazard rate and a covariate vector. This model, operating with continuous time, is often used for sports modelling as well. Many types of sport such as tennis or volleyball, however, are played in a strictly discrete matter with steps divided by individual *points*, *games* or *sets*. Other real life situations with similar properties can be found as well, including the recurrence of diseases or recidivism in crime. Thus authors thus recently presented a novel discrete time random process model - discrete random walk with varying probabilities - which aims to serve as a discrete alternative to the standard Cox's model ZDROJ. In this paper selected properties of the model are presented and possible real life implementations of the model are discussed.

The rest of the paper is organized as follows. Next section.... The last section concludes this work.

2 Random walk with varying probability - overview

The original motivation of the model comes from modelling of sport events where the probability of a success (i.e. scoring a goal, achieving a point etc.) is at the center of interest. After each occurrence of such success this probability either decreases or increases, and thus two basic model alternatives exist - *success punishing* and *success rewarding*. The basic version of the model has 2 parameters, starting success probability p_0 and a memory coefficient λ affecting the severity of probability change after a success. Formally the walk is defined as follows:

Definition 1. Let $p_0, \lambda \in (0, 1)$ be constant parameters, $\{P_n\}_{n=0}^{\infty}$ and $\{X_n\}_{n=1}^{\infty}$ sequences of discrete random variables with $P_0 = p_0$ and for $t \geq 1$

$$P(X_t = 1|P_{t-1} = p_{t-1}) = p_{t-1}, \ \ P(X_t = -1|P_{t-1} = p_{t-1}) = 1 - p_{t-1},$$

and (*success punishing*)

$$P_t = \lambda P_{t-1} + \frac{1}{2}(1-\lambda)(1-X_t) \tag{1}$$

or (*success rewarding*)

$$P_t = \lambda P_{t-1} + \frac{1}{2}(1-\lambda)(1+X_t).$$

The sequence $\{S_n\}_{n=0}^{\infty}$, $S_N = S_0 + \sum_{i=1}^{N} X_i$ for $n \in \mathbb{N}$, with $S_0 \in \mathbb{R}$ some given starting position, is called a *random walk with varying probabilities*, with $\{X_n\}_{n=1}^{\infty}$ being the steps of the walker and $\{P_n\}_{n=0}^{\infty}$ transition probabilities. Depending on the chosen formula to calculate P_i the walk type is either *success punishing* REF or *success rewarding* REF.

The model was first introduced in DDNY and a more thorough description was provided in TEZE. More properties were then presented in AMISTAT and a practical implementation of the model in modelling tennis matches was presented in IMAMAN+ProceedingsAteny. Next let us recall the results of the previous work.

2.1 Success punishing model

The basic properties of the *success punishing* version of the walk are presented in this section. They are presented as a set of expressions, the reader is kindly asked to see referred papers for full proves of those expressions.

For the expected value and variance of the step size for the $t \geq 1$ iteration of the walk X_t it holds AMISTAT

$$EX_t = (2\lambda - 1)^{t-1}(2p_0 - 1),$$

$$Var\, X_t = 1 - (2\lambda - 1)^{2(t-1)}(2p_0 - 1)^2.$$

For the expected value and variance of the transition probability or the $t \geq 1$ iteration of the walk P_t it holds DDNY/TEZE, AMISTAT

$$EP_t = (2\lambda - 1)^t p_0 + \frac{1 - (2\lambda - 1)^t}{2}, \tag{2}$$

$$Var\, P_t = (3\lambda^2 - 2\lambda)^t p_0^2 + \sum_{i=0}^{t-1} K(i)(3\lambda^2 - 2\lambda)^{t-1-i} - k(t)^2,$$

where

$$k(t) = EP_t = (2\lambda - 1)^t p_0 + \frac{1 - (2\lambda - 1)^t}{2}$$

and

$$K(t) = k(t) \cdot (-3\lambda^2 + 4\lambda - 1) + (1 - \lambda)^2.$$

Finally, the expected position of the walker S_t after $t \geq 1$ iterations can be expressed as DDNY/TEZE

$$ES_t = S_0 + (2p_0 - 1)\frac{1 - (2\lambda - 1)^t}{2(1 - \lambda)}. \tag{3}$$

2.2 Success rewarding model

Similar formulas can be derived for the *success rewarding* model. Once again only the formulas are presented with proofs in the referred literature. For the sake of clarity the set of expressions is presented in the same manner as in the previous section.

For the expected value and variance of the step size for the $t \geq 1$ iteration of the walk X_t it holds AMISTAT

$$EX_t = 2p_0 - 1,$$

$$Var\, X_t = 4p_0(1 - p_0).$$

For the expected value and variance of the transition probability or the $t \geq 1$ iteration of the walk P_t it holds AMISTAT

$$EP_t = p_0, \tag{4}$$

$$Var\, P_t = (2\lambda - \lambda^2)^t p_0^2 + p_0(1 - \lambda)^2 \sum_{i=0}^{t-1} (2\lambda - \lambda^2)^i - p_0^2.$$

As the sum in the formula equals $\frac{1-(2\lambda-\lambda^2)^t}{1-(2\lambda-\lambda^2)}$, it can be further simplified as

$$Var\, P_t = (2\lambda - \lambda^2)^t p_0^2 + p_0(1 - \lambda)^2 \frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2} - p_0^2 =$$

$$= p_0[(2\lambda - \lambda^2)^t (p_0 - 1) + 1] - p_0^2,$$

$$Var\, P_t = p_0[(1 - p_0)(1 - (2\lambda - \lambda^2)^t)].$$

Finally, the expected position of the walker S_t after $t \geq 1$ iterations can be expressed as AMISTAT

$$ES_t = S_0 + t(2p_0 - 1). \tag{5}$$

3 Random walk with varying probability - properties

The last formulas missing in the previous section are those expressing the variance of the position of the walker $Var\, S_t$. They will be proved in this section. Let us starting with the *success punishing* model and first prove a support proposition.

Proposition 2. *For all* $t \geq 1$

$$E(P_t S_t) = (2\lambda - 1)^t p_0 S_0 + \sum_{i=0}^{t-1} (2\lambda - 1)^i q(t - 1 - i; p_0, S_0, \lambda), \tag{6}$$

where

$$q(j; p_0, S_0, \lambda) = (1 - \lambda)s(j; p_0, S_0, \lambda) + 2\lambda\pi(j; p_0, \lambda) + (1 - 2\lambda)p(j; p_0, \lambda) + \lambda - 1$$

and $p(j; p_0, \lambda) = EP_j$ *given by (2),* $s(j; p_0, S_0, \lambda) = ES_j$ *given by (3) and* $\pi(j; p_0, \lambda) = EP_j^2$ *given as*

$$EP_t^2 = (3\lambda^2 - 2\lambda)^t p_0^2 + \frac{1 - (3\lambda^2 - 2\lambda)^t}{3\lambda + 1} \frac{(\lambda + 1)}{2} - (p_0 - \frac{1}{2})(3\lambda - 1)(\lambda - 1)M(t - 1; \lambda), \tag{7}$$

where

$$M(t; \lambda) = \sum_{i=0}^{t-1} (3\lambda^2 - 2\lambda)^{t-1-i} (2\lambda - 1)^i.$$

Proof. The formula for EP_t^2 follows from the proof of Proposition 2.5 in AMISTAT. To prove (6) let us start with expressing the value of $E(P_t S_t)$ from the knowledge of past steps as

$$E(P_t S_t) = E[E(P_t S_t | P_{t-1})] = E[E((\lambda P_{t-1} + \frac{1}{2}(1 - \lambda)(1 - X_t))(S_{t-1} + X_t) | P_{t-1})] =$$

$$= E[E(\lambda P_{t-1} S_{t-1} + \frac{1 - \lambda}{2} S_{t-1} - \frac{1 - \lambda}{2} X_t S_{t-1} + \lambda X_t P_{t-1} + \frac{1 - \lambda}{2} X_t - \frac{1 - \lambda}{2} X_t^2 | P_{t-1})]$$

and using $E(X_t | P_{t-1}) = 2P_{t-1} - 1$ and $EX_t^2 = 1$ finally

$$E(P_t S_t) = (2\lambda - 1)E(P_{t-1} S_{t-1}) + (1 - \lambda)ES_{t-1} + 2\lambda EP_{t-1}^2 + (2\lambda - 1)EP_{t-1} + \lambda - 1. \tag{8}$$

Further we will continue using mathematical induction. For $t = 1$ using the definition of the walk it holds that

$$E(P_1 S_1) = p_0(p_0 \lambda(S_0 + 1)) + (1 - p_0)[(1 - (1 - p_0)\lambda)(S_0 - 1)] =$$

$$= (2\lambda - 1)p_0 S_0 + (1 - \lambda)S_0 + 2\lambda p_0^2 - (2\lambda - 1)p_0 + \lambda - 1.$$

When substituting$t = 1$ into (6) we obtain

$$E(P_1S_1) = (2\lambda - 1)p_0S_0 + \sum_{i=0}^{0}(2\lambda - 1)^i q(0 - i; p_0, S_0, \lambda) =$$

$$= (2\lambda - 1)p_0S_0 + (1 - \lambda)s(0; p_0, \lambda) + 2\lambda\pi(0; p_0, \lambda) + (1 - 2\lambda)p(0; p_0, \lambda) + \lambda - 1$$

and finally using (2), (3) and (7)

$$E(P_1S_1) = (2\lambda - 1)p_0S_0 + (1 - \lambda)S_0 + 2\lambda p_0^2 + (1 - 2\lambda)p_0 + \lambda - 1.$$

Equation (6) thus holds $t = 1$. Now for the induction step $t \to t + 1$ we get by substituting(6) into (8)

$$E(P_{t+1}S_{t+1}) = (2\lambda - 1)E(P_tS_t) + (1 - \lambda)ES_t + 2\lambda EP_t^2 + (2\lambda - 1)EP_t + \lambda - 1 =$$

$$(2\lambda-1)[(2\lambda-1)^t p_0S_0 + \sum_{i=0}^{t-1}(2\lambda-1)^i q(t-1-i)] + (1-\lambda)s(t) + 2\lambda\pi(t) + (2\lambda-1)p(t) + \lambda - 1 =$$

$$(2\lambda - 1)^{t+1}p_0S_0 + \sum_{i=0}^{t}(2\lambda - 1)^i q(t - i).$$

□

Theorem 3. *For all $t \geq 1$*

$$Var\, S_t = t(1 - 2S_0) + 4\sum_{i=0}^{t-1}\sigma(i; p_0, S_0, \lambda) - a(t; p_0, S_0, \lambda),$$

with $\sigma(i; p_0, S_0, \lambda) = E(P_tS_t)$ given by (6) and

$$a(t; p_0, S_0, \lambda) = (2p_0 - 1)\sum_{i=0}^{t-1}\frac{1 - (2\lambda - 1)^i}{1 - \lambda} + S_0(2p_0 - 1)\frac{1 - (2\lambda - 1)^t}{1 - \lambda} + (2p_0 - 1)^2\frac{(1 - (2\lambda - 1)^t)^2}{4(1 - \lambda)^2}.$$

Proof. From the definition of variance

$$Var\, S_t = ES_t^2 - E^2S_t \tag{9}$$

and (3) follows that to prove the proposition it is enough to prove that

$$ES_t^2 = S_0^2 + t(1 - 2S_0) + 4\sum_{i=0}^{t-1}\sigma(i; p_0, S_0, \lambda) - (2p_0 - 1)\sum_{i=0}^{t-1}\frac{1 - (2\lambda - 1)^i}{1 - \lambda}. \tag{10}$$

First of all, let us express ES_t^2 from the knowledge of past walk development. From the definition of the expected value and the definition of the walk (Definition 1) it follows

$$ES_t^2 = E[E(S_t^2|P_{t-1})] = E[E((S_{t-1} + X_t)^2|P_{t-1})] = E(S_{t-1}^2 + 2(2P_{t-1} - 1)S_{t-1} + 1) =$$

$$= ES_{t-1}^2 + 4E(P_{t-1}S_{t-1}) - 2ES_{t-1} + 1, \tag{11}$$

where the fact that $EX_t^2 = 1$ was used. The theorem will be once again proved using mathematical induction. For $t = 1$ we get using the definition of the walk (Def (1))

$$ES_1^2 = p_0(S_0+1)^2 + (1-p_0)(S_0-1)^2 =$$
$$= S_0^2 - 2S_0 + 4p_0S_0 + 1.$$

Substituting $t = 1$ into (10) we obtain

$$ES_1^2 = S_0^2 + 1 - 2S_0 + 4\sigma(0; p_0, S_0, \lambda) - (2p_0 - 1)\frac{1-(2\lambda-1)^0}{1-\lambda} =$$
$$= S_0^2 + 1 - 2S_0 + 4p_0S_0$$

and (10) thus holds for $t = 1$. Now for the induction step $t \to t+1$ we get by substituting (10) into (11)

$$ES_{t+1}^2 = ES_t^2 + 4E(P_tS_t) - 2ES_t + 1 =$$
$$= S_0^2 + t(1-2S_0) + 4\sum_{i=0}^{t-1}\sigma(i; p_0, S_0, \lambda) - (2p_0-1)\sum_{i=0}^{t-1}\frac{1-(2\lambda-1)^i}{1-\lambda} +$$
$$+4\sigma(t; p_0, S_0, \lambda) - 2(S_0 + (2p_0-1)\frac{1-(2\lambda-1)^t}{2(1-\lambda)}) + 1 =$$
$$S_0^2 + (t+1)(1-2S_0) + 4\sum_{i=0}^{t}\sigma(i; p_0, S_0, \lambda) - (2p_0-1)\sum_{i=0}^{t}\frac{1-(2\lambda-1)^i}{1-\lambda}.$$

Substituting (10) and (3) into (9) then proves the theorem. □

For the *success rewarding* model is the procedure similar. First let us prove a support proposition.

Proposition 4. *For all* $t \geq 1$

$$E(P_tS_t) = p_0S_0 + p_0t + 2\lambda p_0(p_0-1)\frac{1-(2\lambda-\lambda^2)^t}{(1-\lambda)^2}. \tag{12}$$

Proof. We will once again start with expressing $E(P_tS_t)$ from the knowledge of the past step.

$$E(P_tS_t) = E(E(P_{t-1}S_{t-1}|P_{t-1}) = E[E((\lambda P_{t-1} + \frac{1}{2}(1-\lambda)(1+X_t))(S_{t-1}+X_t)|P_{t-1})] =$$
$$= E[E(\lambda P_{t-1}S_{t-1} + \frac{1-\lambda}{2}S_{t-1} + \frac{1-\lambda}{2}X_tS_{t-1} + \lambda X_tP_{t-1} + \frac{1-\lambda}{2}X_t + \frac{1-\lambda}{2}X_t^2|P_{t-1})]$$

and using $E(X_t|P_{t-1}) = 2P_{t-1} - 1$ and $EX_t^2 = 1$ finally

$$E(P_tS_t) = E(P_{t-1}S_{t-1}) + 2\lambda EP_{t-1}^2 - (2\lambda-1)EP_{t-1}. \tag{13}$$

Further we will continue using mathematical induction. For $t = 1$ using the definition of the walk it holds that

$$E(P_1 S_1) = p_0(1 - (1 - p_0)\lambda)(S_0 + 1) + (1 - p_0)\lambda p_0(S_0 - 1) =$$

$$= p_0 S_0 + 2\lambda p_0^2 - (2\lambda - 1)p_0.$$

When substituting$t = 1$ into (12) we obtain

$$E(P_1 S_1) = p_0 S_0 + p_0 + 2\lambda p_0(p_0 - 1)\frac{1 - (2\lambda - \lambda^2)^0}{(1 - \lambda)^2} =$$

$$= p_0 S_0 + p_0 + 2\lambda p_0(p_0 - 1)$$

and finally

$$E(P_1 S_1) = p_0 S_0 + 2\lambda p_0^2 - (2\lambda - 1)p_0.$$

Equation (12) thus holds $t = 1$. Now for the induction step $t \to t + 1$ we get by substituting(12) into (13)

$$E(P_{t+1} S_{t+1}) = E(P_t S_t) + 2\lambda E P_t^2 - (2\lambda - 1)E P_t$$

and further using

$$E P_t^2 = p_0((2\lambda - \lambda^2)^t (p_0 - 1) + 1),$$

which follows from the proof of Proposition 3.7 in AMISTAT, and (4)

$$E(P_{t+1} S_{t+1}) = p_0 S_0 + p_0 t + 2\lambda p_0(p_0 - 1)\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2} + 2\lambda p_0((2\lambda - \lambda^2)^t (p_0 - 1) + 1) - (2\lambda - 1)p_0 =$$

$$= p_0 S_0 + p_0(t + 1) + 2\lambda p_0(p_0 - 1)[\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2} + (2\lambda - \lambda^2)^t] =$$

$$= p_0 S_0 + p_0(t + 1) + 2\lambda p_0(p_0 - 1)\frac{1 - (2\lambda - \lambda^2)^{t+1}}{(1 - \lambda)^2}.$$

□

Theorem 5. *For all $t \geq 1$ holds*

$$Var\, S_t = 4p_0(1 - p_0)t^2 + a(p_0, \lambda)t - a(p_0, \lambda)\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2},$$

where

$$a(p_0, \lambda) = \frac{8p_0(1 - p_0)}{(1 - \lambda)^2}.$$

Proof. As clearly the value S_0 does not affect the variance, let us from now assume $S_0 = 0$. From the definition of variance and (5) follows that to prove the theorem it is enough to prove that

$$E S_t^2 = t^2 + a(p_0, \lambda)t - a(p_0, \lambda)\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2}. \tag{14}$$

First of all let us recall that formula (11) holds for the *success rewarding* type of the model as well. The theorem will be once again proved using mathematical induction. For $t = 1$ the definition of the walk yields the same result as in the proof of Theorem (3). Substituting $t = 1$ into (14) we obtain

$$ES_1^2 = 1 + a(p_0, \lambda)t - a(p_0, \lambda) = 1$$

and (14) thus holds for $t = 1$. Now for the induction step $t \to t + 1$ we get by substituting (14), (12) and (5) into (11)

$$ES_{t+1}^2 = ES_t^2 + 4E(P_t S_t) - 2ES_t + 1 =$$

$$= t^2 + a(p_0, \lambda)t - a(p_0, \lambda)\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2} + 4(p_0 t + 2\lambda p_0(p_0 - 1)\frac{1 - (2\lambda - \lambda^2)^t}{(1 - \lambda)^2}) - 2t(2p_0 - 1) + 1 =$$

$$= (t + 1)^2 + a(p_0, \lambda)(t + 1) - a(p_0, \lambda)\frac{1 - (2\lambda - \lambda^2)^{t+1}}{(1 - \lambda)^2}.$$

Substituting (14) and (5) into the definition of variance then proves the theorem.

□

4 Conclusion