Discrete random walks with memory: Models and applications

Tomáš Kouřim

November 21, 2019

Abstract

The contribution focuses on non-Markov random processes, i.e. processes with memory, and is especially concerned with random processes where either one or just a small number of events significantly affects its future development. Reliability analysis, medical studies or sport statistics provide many real-life examples of such processes. This paper focuses on statistical event-history analysis with discretized time, where the role of the hazard rate can be substituted by different variants of transition probabilities. In the simplest case the Bernoulli scheme is used.

The main concern of the paper is the formulation of models describing the dependence of transition probabilities on the process development, as well as on exogenous factors – covariates. Such an impact can be incorporated explicitly and transition probabilities modulated using a few parameters reflecting the current state of the walk as well as the information about the past path. In more complicated cases, as well as in the presence of exogenous covariates, the changes of probabilities are modeled via a regression model, for instance the logistic one. The behavior of proposed random walks is studied both theoretically and with the aid of simulations. Finally, the approach is illustrated on several real data examples.

1 Introduction

Random wak is a well described mathematicall object first introduced by K.Pearson in 1905 [3]. Since then many variations of a random walk have been derived and described and it was applied on many reali life problems (zdroje). Yet there are still new posibilities and options how to alter and improve the classical random walk and present yet another new model representing different real life events. One of such modifications is the random walk with varying step size indroduced in 2010 by Turban [4] which served as an inspiration to the random walk with varying probabilities introduced by Kouřim [1, 2]. It was later shown that such a random walk can be sucesfully applied to predict *in-play* odds and used for actual betting against one of the comercial bookmakers (zdroj Ateny, work in progress). In this paper, the theoretical properties of the model are described and furher examined and the results are tested on generated data. The rest of the paper is organized as follows....

2 Random walk with varying probabilities

The random walk with varying probabilities is based on a standard Bernoulli (zdroj) random walk with some starting transition probability p_0. This probability is then altered after each step of the walk using a coefficient λ so that the repetition of the same step becomes less probable. Formally, it can be defined [2]

Definition 1. Let $\{X_n\}_{n=1}^{\infty}$ and $\{P_n\}_{n=1}^{\infty}$ be sequences of discrete random variables, and $p_0 \in [0, 1]$ and $\lambda \in (0, 1)$ constant parameters, such that the first random variable X_1 is given by

$$P(X_1 = 1) = p_0$$

$$P(X_1 = -1) = 1 - p_0.$$

Further

$$P_1 = \lambda p_0 + \frac{1}{2}(1-\lambda)(1 - X_1) \tag{1}$$

and for $i \geq 2$

$$P(X_i = 1 | P_{i-1} = p_{i-1}) = p_{i-1}$$

$$P(X_i = -1 | P_{i-1} = p_{i-1}) = 1 - p_{i-1}$$

$$P_i = \lambda P_{i-1} + \frac{1}{2}(1-\lambda)(1 - X_i). \tag{2}$$

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 $\{P_n\}_{n=1}^{\infty}$ transition probabilities.

From [2], it can be further derived that at each step $t + k$, $t, k > 0$ the value of a transition probability P_{t+k} can be computed from the knowledge of transition probability P_t and the realization of the walk $X_{t+1}, \ldots, X_{t+k}$ using formula

$$P_{t+k} = P_t \lambda^{t+k} + \frac{1}{2}(1-\lambda) \sum_{i=t+1}^{t+k} \lambda^{t+k-i}(1 - X_i).$$

2.1 Properties

Basic properties of the random walk with varying are described in [2], namely that

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

and

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

for $\forall t \geq 1$. This further yields $EP_t \to \frac{1}{2}$ and $ES_t \to S_0 + \frac{2p_0-1}{2(1-\lambda)}$ for $t \to +\infty$.

Now to describe the variance of the transition probability, let us proove the following propositions.

Proposition 1. *For $\forall t \geq 1$, it holds that*

$$Var(P_t) = TODO$$

Proof. TODO

And similarly for the variance of the position of the walker, following statements can be proved. □

Proposition 2. *For $\forall t \geq 1$, it holds that*

$$Var(S_t) = TODO$$

Proof. TODO □

3 Random walk with varying transition probability - alternatives

3.1 Success rewarded

The basic definition of the random walk (1) presents a model, where the "success is punished", meaning the probability of an event is decreased every time that event occurs. Opposite situation can be considered, where the probability of an event is increased every time that event occurs. Formally, such a random walk is defined in a following manner [2].

Definition 2. Let $\{X_n\}_{n=1}^{\infty}$, p_0 and λ be as in Definition 1. Further let $\{P_n\}_{n=1}^{\infty}$ be a sequence of discrete random variables given by

$$P_1 = \lambda p_0 + \frac{1}{2}(1-\lambda)(1+X_1) \tag{3}$$

$$P_i = \lambda P_{i-1} + \frac{1}{2}(1-\lambda)(1+X_i) \ \ \forall i \geq 2. \tag{4}$$

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 a random walk with varying probabilities.

This version of the random walk presents a model where "success is rewarded". For the sake of clarity let us denote all variables connected with the "success punished" model from Definition 1 with the subscript P (as ${}_PP$, ${}_PS$) and the "success rewarded" version from Definition 2 with subscript R (as ${}_RP$, ${}_RS$). For the reward version it is easy to prove similar proposition as for the punish version.

Proposition 3. *For* $\forall t \geq 2$,

$$ {}_R P_t = p_0 \lambda^t + \frac{1}{2}(1-\lambda) \sum_{i=1}^{t} \lambda^{t-i}(1+X_i) \tag{5} $$

Proof. The proposition is proved using induction. For $t = 2$ using 1 it holds that

$$ {}_R P_2 = \lambda {}_R P_1 + \frac{1}{2}(1-\lambda)(1+{}_R X_2) = \lambda(\lambda p_0 + \frac{1}{2}(1-\lambda)(1+X_1)) + \frac{1}{2}(1-\lambda)(1+{}_R X_2) = $$

$$ = p_0 \lambda^2 + \frac{1}{2}(1-\lambda) \sum_{i=1}^{2} \lambda^{2-i}(1+X_i), $$

which is in accordance with 5. Now for the induction step $t \to t+1$ we obtain from 2 and the induction assumption

$$ {}_R P_{t+1} = \lambda {}_R P_t + \frac{1}{2}(1-\lambda)(1+{}_R X_{t+1}) = \lambda(p_0 \lambda^t + \frac{1}{2}(1-\lambda) \sum_{i=1}^{t} \lambda^{t-i}(1+X_i)) + \frac{1}{2}(1-\lambda)(1+{}_R X_{t+1}) = $$

$$ = p_0 \lambda^{t+1} + \frac{1}{2}(1-\lambda) \sum_{i=1}^{t} \lambda^{t-i+1}(1+X_i) + \frac{1}{2}(1-\lambda)(1 +_R X_{t+1}) = $$

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

$\square$

Proposition 4. *For* $\forall t \geq 1$, $E({}_R P_t) = \lambda_0$.

Proof. Using $E({}_R X_t | {}_R P_{t-1}) = 2 {}_R P_{t-1} - 1$ and 4 we obtain

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

$$ = E[\lambda {}_R P_{t-1} + \frac{1}{2}(1-\lambda)(1 + 2 {}_R P_{t-1} - 1)] = E[\lambda {}_R P_{t-1} + (1-\lambda) {}_R P_{t-1}) = $$

$$ = E_R P_{t-1}. $$

Recursively we get

$$ E_R P_t = E p_0 = p_0. $$

$\square$

Now to calculate the position of the walker at a given step $t \geq 1$, it is easy to see that ${}_R S_t = {}_R S_{t-1} + 2 {}_R P_{t-1} - 1$. From this, we can prove the following statement about the expected position of the walker.

Proposition 5. *For* $\forall t \geq 1$,

$$E({}_R S_t) ={}_R S_0 + t(2p_0 - 1).$$

Proof. Using the result of Proposition 4 we get

$$E({}_R S_{t+1}) = E[E({}_R S_{t+1}|{}_R S_t)] = E[{}_R S_t + E(2{}_R P_{t-1} - 1)] =$$
$$= E_R S_t + (2p_0 - 1)]$$

which recursively proves the statement. □

Proposition 6. *For* $\forall t \geq 1$, $Var({}_R P_t) = 0$.

Proof. Bych tak tipoval □

Proposition 7. *For* $\forall t \geq 1$, $Var({}_R S_t) = 0$.

Proof. Tipuju

^^tohle jsou vsechno nove veci, v tezich se o nich jenom zminuju, ale nedavam dukazy. Mozna by se to mohlo nejakym zpusobem sjednotit. Treba vyrknout vetu, ktera by rikala tohle vsechno, a pak to dokazat naraz. □

vytvorit nove simulace a obrazky ukazani jednotlivych hodnot, protovnani - pozor na floating point zaokrouhlovani

3.2 Two lambdas

Another level of complexity can be added by using separate λ parameters for each direction of the walk. Again, two ways of handling success are available. The success punished version is defined as follows.

Definition 3. Let $\{X_n\}_{n=1}^{\infty}$ and p_0 be as in Definition 1. Further let λ_0, $\lambda_1 \in (0,\,1)$ be constant coefficients and $\{P_n\}_{n=1}^{\infty}$ be a sequence of discrete random variables given by

$$P_1 = \frac{1}{2}[(1 + X_1)\lambda_0 p_0 + (1 - X_1)(1 - \lambda_1(1 - p_0))] \tag{6}$$

$$P_i = \frac{1}{2}[(1 + X_i)\lambda_0 P_{i-1} + (1 - X_i)(1 - \lambda_1(1 - P_{i-1}))] \ \ \forall i \geq 2. \tag{7}$$

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 a random walk with varying probabilities.

And the success rewarded version as

Definition 4. Let $\{X_n\}_{n=1}^{\infty}$ and p_0 be as in Definition 1. Further let λ_0, $\lambda_1 \in (0,\,1)$ be constant coefficients and $\{P_n\}_{n=1}^{\infty}$ be a sequence of discrete random variables given by

$$P_1 = \frac{1}{2}[(1 - X_1)\lambda_0 p_0 + (1 + X_1)(1 - \lambda_1(1 - p_0))]$$

$$P_i = \frac{1}{2}[(1 - X_i)\lambda_0 P_{i-1} + (1 + X_i)(1 - \lambda_1(1 - P_{i-1}))] \;\; \forall i \geq 2.$$

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 a random walk with varying probabilities.

TODO mel bych to rovnou nejak znacit $^1_R P$ a $^2_S P$ jako reward s 1 lambda a success s 2 lambda? Abych se zas neuindexoval

Let us prove the following propositions describing the properties of a random walk affected by two coefficients lambda.

TODO stejne veci opet dokazat pro tuhle variantu

3.3 Other alternatives

The presented model of a random walk can be further developed and more versions can be derived and described. These variants include but are not limited to multidimensional walk (with either one or multiple λ parameters, again with *success rewarded* or *success punished*), a walk with the transition probability explicitly dependent on more than the last step, i.e. $P_t(k) \sim P_t(X_t, X_{t-1}, \ldots, X_{t-(k-1)})$, or the walk with λ parameter not constant, but a function of the time t, i.e. $P_t(\lambda(t))$. Detailed properties of such walks together with their possible applications on real life problems will by subject of a further study.

4 Simulations

Nasimulovani ruznych druhu prochazky s ruznymi parametry. Nasledne pokus o zpetne odhaleni druhu prochazky a jejich paramteru. Budu to delat zrejme podle MLE, vyhodnocovat asi podle te nejvetsi verohodnosti, pripadne podle goodness-of-fit

5 Results

Zhodnoceni prace, ze je to vlastne skvele, ze neco takoveho je, a na co vsechno by se to vlastne dalo pouzit.

References

[1] Tomáš Kouřim. Random walks with varying transition probabilities. *Doktorandské dny FJFI*, 2017. Available at http://kmwww.fjfi.cvut.cz/ddny/historie/17-sbornik.pdf.

[2] Tomáš Kouřim. Statistical analysis, modeling and applications of random processes with memory. *PhD Thesis Study, ČVUT FJFI*, 2019.

[3] Karl Pearson. The problem of the random walk. *Nature*, 72(1865):294, 1905.

[4] Loïc Turban. On a random walk with memory and its relation with markovian processes. *Journal of Physics A: Mathematical and Theoretical*, 43(28):285006, 2010.