Physica A 406 (2014) 230–235

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

An exactly solvable correlated stochastic process in finite time

Jongwook Kim [a,*], Junghyo Jo [a,b]

[a] Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea
[b] Department of Physics, POSTECH, Pohang, Republic of Korea

HIGHLIGHTS

- A correlated stochastic process inspired by the Ehrenfest urn is proposed.
- Moment generating function of the model is exactly obtained.
- Correlation in financial time-series data is inferred by using the model.

ARTICLE INFO

Article history:
Received 17 January 2014
Received in revised form 20 February 2014
Available online 21 March 2014

Keywords:
Ehrenfest urn
Time-series data
Bayesian statistics

ABSTRACT

We propose a correlated stochastic process of which the novel non-Gaussian probability mass function is constructed by exactly solving moment generating function. The calculation of cumulants and auto-correlation shows that the process is convergent and scale invariant in the large but finite time limit. We demonstrate that the model infers the correlation strength in a discrete correlated time-series data, and predicts the data distribution with high precision in the finite time regime.

1. Introduction

Non-normal distributions and clustering are prevalent in physical, social and biological phenomena, such as in crystal growth, polymer transportation/distribution [1,2], brain electrical activity [3], financial time-series data [4,5], social networks [6], and extremely rare events [7,8]. In particular, heavy-tail and clustering are frequently observed among them. Leptokurtic distributions have been modeled by a family of stable distributions, constructed from the composite of independent and nonidentical Brownian and Poisson-jump processes. Clustering in time-series events is caused by the correlations between these events, for which various continuous stochastic models have been proposed to explain their phenomena. Auto-regressive moving average models [5] are popular continuous Gaussian stochastic models with correlated properties. Fractional Brownian motion [9] is a continuous stochastic process, whose model is defined by the exponent of the power-law auto-correlation function of two white noises at different times.

There are also various discrete models of correlated random walks, collectively referred to as urn models. Classical urn schemes are Polya's model [10], which is implied by the β-distribution, and Friedman's model [11], which is a kind of dual to Polya's model. However, not much attention has been given to urn models in the analysis of time-series events, compared to the use of various continuous volatility models. Due to the inflexibility and limitations of current stochastic modelings, there is no unified prescription for the analysis of all kinds of data, and only a few discrete processes are solved analytically.

* Corresponding author.
E-mail addresses: dr.jongwookkim@gmail.com (J. Kim), jojunghyo@gmail.com, jojunghyo@apctp.org (J. Jo).

http://dx.doi.org/10.1016/j.physa.2014.03.055

In such circumstances, it is worth developing a new discrete stochastic scheme constructed from a discrete micro process to explain the non-normality of time-series data using correlations.

This paper consists of four sections. In Section 2, we introduce a generalized Ehrenfest urn model and obtain the exact solution of the moment-generating function for the correlated stochastic process. Then, we apply our model to financial data in Section 3, and summarizes our results in Section 4.

2. Model

Polya's urn contains balls of two colors. Drawing a ball randomly, one puts an additional ball of the same color into the urn, as well the originally drawn ball. Thus dominant colored balls become more dominant because positive correlations exist between drawing events. On the other hand, Ehrenfest introduced a different process concerning two urns, originally to explain the dissipation of heat. Initially, one urn is filled with $\mathcal{N}$ labeled balls, while the other urn is empty. In each time step, one randomly chooses a number from 1 to $\mathcal{N}$, and moves the ball corresponding to that number into the opposite urn. Therefore, the serial events are negatively correlated proportional to $-1/\mathcal{N}$.

In this paper, we introduce a generalized Ehrenfest urn scheme to model not only negatively but also positively correlated stochastic processes within finite time intervals. Note that Polya's urn also describes positive correlations between events, but total ball number increases with time unlike the Ehrenfest urn scheme. The model has two urns (Urn^+ and Urn^-) with $2\mathcal{N}$ labeled balls. Assuming unskewedness, we initially put $\mathcal{N}$ balls in each urn. At each time, one randomly chooses an integer number between 1 and $2\mathcal{N}$. To incorporate *negative* correlation between events as in the original Ehrenfest model, when the ball of the chosen number was in Urn^+, we move a ball in Urn^+ into Urn^-. On the other hand, to incorporate *positive* correlation, when the ball of the chosen number was in Urn^+, we move a ball in Urn^- into Urn^+.

In this scheme, the direction of ball movement is opposite to the original Ehrenfest model. After one round of drawing, one has $\mathcal{N}+1$ balls in Urn^+ and $\mathcal{N}-1$ balls in Urn^-; therefore, the probability of drawing a ball in Urn^+ in the second round is $1/2+1/2\mathcal{N}$. Successive drawing events are positively correlated, and the correlation strength ϵ is $1/2\mathcal{N}$. If we allow to move multiple (e.g., $\mathcal{M}$) balls at each drawing, the correlation strength ϵ can be increased as $\mathcal{M}/2\mathcal{N}$ in general.

In the stochastic process, we have a particular interest in finite iteration time N. Here the iteration time N is sufficiently small compared with the total ball number $\mathcal{N}$. Otherwise, the process with the positive correlation can become ill-defined after many iterations when one urn is empty so that it cannot provide balls to the opposite urn. Although the original Ehrenfest process has been solved only for the infinite iterations [10,12], it has not been analytically solved for the finite iteration to our knowledge.

Before formulating the correlated stochastic process, it is necessary to introduce a *genuine* correlation

$$\kappa = \epsilon N \tag{1}$$

that rescales the correlation strength ϵ by multiplying the total iteration time N. It may not be trivial that the stochastic model generates the same stationary distribution for short iterations with strong correlation versus for long iterations with weak correlation.

Now we define a stochastic position $m_n = \sum_{i=1}^{n} \delta m_i$ at the nth iteration, where the unit-sized displacement δm_i at the ith time is either $+1$ or -1 depending on the outcome of the ball drawn from Urn^+ or Urn^-. Note that the stochastic position indicates half of the ball number difference between the two urns. This corresponds to a modified binomial process with the presence of correlation between events. The correlated stochastic process can also be described by a recursive master equation for the probability mass function(PMF), $P(m, n)$, with a transition rate modified from the binomial process:

$$P(m, n+1) = \left[\frac{1}{2} + (m-1)\epsilon\right]P(m-1, n) + \left[\frac{1}{2} - (m+1)\epsilon\right]P(m+1, n), \tag{2}$$

where $n = \{0, 1, 2, \ldots, N\}$ and its initial PMF is $P(0, 0) = 1$. The discrete location m_n, realized at the time after n, runs from $-n$ to n in steps of 2, i.e., $m_n \in \{-n, -n+2, -n+4, \ldots, n-2, n\}$. In the continuous limit of position ($x = a \cdot m$) and time ($t = n \cdot \tau$) with infinitesimal increments a and τ, respectively, the process converges to the Ornstein–Uhlenbeck-like distribution [13]: $\partial_t p(x, t) = -2(a\epsilon/\tau)\partial_x[xp(x, t)] + 1/2(a^2/\tau)\partial_x^2 p(x, t)$. Negative ϵ gives a stationary Gaussian distribution of $p(x, t)$, while positive ϵ results in the sign of drift term to be opposite to that of the conventional Ornstein–Uhlenbeck process, leading to non-Gaussian distributions [14].

After a few hundred iterations, this process describes the conspicuous non-Gaussian properties of finite N statistics. Herein, we demonstrate the utility of our model in this regime by testing it on high-frequency financial time-series data, where the large standard deviation in time and positive auto-correlation are observed as their non-Gaussian entities. As N grows larger, the process becomes time scale invariant so that the cumulants start to converge.

The moment generating function is introduced as $Z_n(q) = \sum_m q^m P(m, n)$, where $Z_n(0) = \sum_m P(m, n) = 1$. The moment generating function at time 0 is defined as $Z_0(q) = 1$, and then the recurrence in Eq. (2) is recast into a differential equation for q, shown as

$$Z_{n+1}(q) = \frac{1}{2}(q + q^{-1})Z_n(q) + \epsilon(q^2 - 1)\partial_q Z_n(q). \tag{3}$$