MODELING AND ESTIMATION OF INTERACTIONS OF YULE-SIMON PROCESSES

Lingqing Gan, Asher A. Hensley, and Petar M. Djurić

Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11790

ABSTRACT

Yule-Simon processes are preferential attachment processes that can be represented as urn processes where balls are added to a growing number of urns and where a new ball is placed in an urn with probability that is proportional to the number of balls in the urn. In this paper, we consider Yule-Simon processes that interact with each other. We propose two models of interaction, a deterministic and a probabilistic model. Based on observed two processes, we want to determine if the two processes interact and the direction of the interaction. In the case of the probabilistic model, the objective is also to estimate the strength of the interaction. We present detection/estimation schemes for each model. We also provide simulation results that demonstrate the performance of our schemes.

Index Terms— Yule-Simon distribution, interaction, detection, estimation, Gibbs sampling

1. INTRODUCTION

Yule-Simon (YS) processes are preferential attachment processes and they give rise to YS distributions, which are realizations of Zipf's law [1]. YS distributions exhibit heavy tail properties and they find a wide range of applications [2, 3]. YS processes can be used in modeling the switching of regimes of time series, where the probability of the time series to switch a regime is proportional to the length of the regime. These processes are often used to model phenomena with bursty dynamics, e.g., in communication systems [4], natural disasters [5], and financial markets [6].

The YS distribution was discovered independently by Yule in 1925 [7] and Simon in 1955 [8] to explain inequality of outcomes in evolution and word frequency, respectively. The basic idea is that selection creates a feedback loop which increases the chances of future selection. This concept has gone under several names such as preferential attachment [9], cumulative advantage [10], and the Matthew effect [11], although there are some technical differences in the latter case.

In this paper we are interested in detecting and estimating interactions between YS processes. We observe two processes and one of them may be affecting the other in an unknown way. For example, if one of the processes undergoes a change of regime, this change may speed up a change of a regime in the other process or vice versa. In our work, we model an influencing process by having the influenced process change its regime more likely than if the first process did not undergo any changes.

We propose two interactive Yule-Simon models. By the first model, $\mathcal{M}_d$, the influencing process affects the other process in a deterministic manner. According to the second model $\mathcal{M}_s$, the influence takes place with a probability p. For model $\mathcal{M}_d$, we propose a Gibbs sampling-based scheme for estimating the parameters of the processes and a method for detecting the influence and its direction. For the $\mathcal{M}_s$ model, we propose a trace-based scheme to calculate the a posteriori distribution of the unknown parameters of the model. We demonstrate the performance of the schemes by simulations.

The main contributions of the paper are the proposed models for interactions between processes and the methods that identify the influences and estimate the unknowns of the system of two processes.

The paper is structured as follows. In Section 2 we birefly introduce the Yule-Simon model. In Section 3 we introduce the $\mathcal{M}_d$ model and the scheme for detecting interactions. In Section 4 we describe the $\mathcal{M}_s$ model and explain the method we propose for obtaining the a posteriori distribution of the parameters of the model. In Section 5, we present the performance of the two schemes from simulated data. With Section 6, we conclude the paper.

2. THE YULE-SIMON MODEL

The YS distribution can be formally expressed in terms of the Beta function and parameter ρ as follows:

$$f(k;\rho) = \rho B(\rho+1,k), \quad k = 1,2,3,..., \tag{1}$$

where for large k, $f(k;\rho) \propto k^{-(\rho+1)}$, thus suggesting that the parameter ρ can be interpreted as the power-law exponent.

We now formulate a temporal stochastic process where the time between events follows the YS distribution [12]. This idea can be represented as a birth-reset process as shown in Figure 1. The circles with numbers represent the duration that the process has been in a given regime. The arrows pointing to the right show the possibility of incrementing the duration of

This work has been supported by NSF under Awards CCF-1618999.

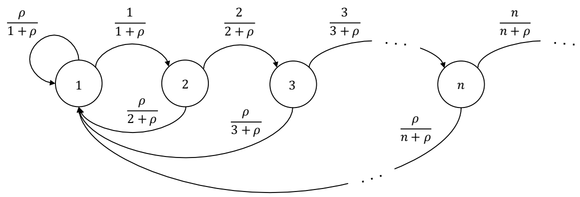

Fig. 1. Temporal representation of the Yule-Simon process.

the regime by one time step with a corresponding probability. The arrows pointing to the left indicate possibility of starting a new regime. As the duration of a regime increases, the probability of starting a new regime decreases, or equivalently, the probability of staying in the same regime increases. The probability of a regime of length k can be written as the product

$$f(k;\rho) = \frac{\rho}{k+\rho}\prod_{j=1}^{k-1}\frac{j}{j+\rho}, \tag{2}$$

which can be shown to be equivalent to (1).

3. A DETERMINISTIC MODEL OF INTERACTIONS

3.1. Problem Statement

Here we observe two YS processes that may interact with each other. We denote them by A and B, and we assume that they operate under one of the following models:

$$\mathcal{M}_0: \ A \text{ does not influence } B,$$
$$\mathcal{M}_1: \ A \text{ influences } B.$$

Under $\mathcal{M}_0$, A and B evolve independently according to the standard YS process described in Section 2. Under $\mathcal{M}_1$, if a reset occurs in A, a reset simultaneously occurs in B. The problem we address is to determine from the observed processes which hypothesis is in effect. Next, we describe the generative models of the processes.

3.2. Generative Models

Define the processes A and B until time instant t as binary time series given by $O_A^{1:t}$ and $O_B^{1:t}$. For each discrete time instant $1 \le \tau \le t$ and process $i \in \{A, B\}$, we have $O_i^\tau \in \{0,1\}$, where $O_i^\tau = 1$ indicates a change of regime, i.e., a reset in the corresponding temporal YS process ($O_i^\tau = 0$, otherwise).

Under $\mathcal{M}_0$, the generative model of the observation at time t can be formulated as follows:

$\mathcal{M}_0$:

$$O_A^t \sim \text{Ber}\left(\frac{s_A^t}{s_A^t + \rho_A}\right), \tag{3}$$

$$O_B^t \sim \text{Ber}\left(\frac{s_B^t}{s_B^t + \rho_B}\right), \tag{4}$$

where $\text{Ber}(q)$ stands for a Bernoulli distribution with probability q, and s_i^t, $i \in \{A, B\}$, is a count used for determining the probability for a regime change (as shown in Fig. 1). At time t, this count is obtained by

$$s_i^t = (s_i^{t-1} + 1)^{1-O_i^t} \quad t = 1, 2 \cdots . \tag{5}$$

Under $\mathcal{M}_1$, the generative model of the observations at time instant t is the same as that under $\mathcal{M}_0$, except that s_A^t and s_B^t are modeled by

$$s_A^t = (s_A^{t-1} + 1)^{1-O_A^t}, \tag{6}$$
$$s_B^t = (s_B^{t-1} + 1)^{(1-O_A^t)(1-O_B^t)}. \tag{7}$$

Thus, the process B will reset its count if process A or process B encounter a regime change. It is important to note that $O_B^t = 1$ if the process B selects a change of regime based on its own probability for switching to a new regime.

3.3. Model Selection

Given sequences of observations from processes A and B, we want to determine which of the two models is correct. We assume that we do not know the parameters ρ_A and ρ_B.

We use a Gibbs sampling algorithm to estimate ρ_A and ρ_B, and it is presented in subsection 3.5. We then use the resulting estimates to calculate the posterior probabilities of each model hypothesis as follows:

$$\begin{aligned} & p(\mathcal{M}_i \mid O_A^{1:t}, O_B^{1:t}) \\ \propto \ & p(O_A^{1:t}, O_B^{1:t} \mid \mathcal{M}_i) p(\mathcal{M}_i). \end{aligned} \tag{8}$$

Further we have,

$$p(O_A^{1:t}, O_B^{1:t} \mid \mathcal{M}_0) = p(O_A^{1:t} \mid \mathcal{M}_0) p(O_B^{1:t} \mid \mathcal{M}_0), \tag{9}$$
$$p(O_A^{1:t}, O_B^{1:t} \mid \mathcal{M}_1) = p(O_B^{1:t} \mid O_A^{1:t}, \mathcal{M}_1) p(O_A^{1:t} \mid \mathcal{M}_1). \tag{10}$$

Model selection then proceeds by applying the MAP rule:

$$\frac{p(\mathcal{M}_0 \mid O_A^{1:t}, O_B^{1:t})}{p(\mathcal{M}_1 \mid O_A^{1:t}, O_B^{1:t})} \gtrless 1, \tag{11}$$

where for our experiments we use a non-informative prior (i.e. $p(\mathcal{M}_0) = p(\mathcal{M}_1) = 0.5$). Since process A is not influenced by process B in neither $\mathcal{M}_0$ nor $\mathcal{M}_1$, after combining (8)-(11), the MAP decision (11) simplifies to

$$\frac{p(O_B^{1:t} \mid O_A^{1:t}, \mathcal{M}_0)}{p(O_B^{1:t} \mid O_A^{1:t}, \mathcal{M}_1)} \gtrless 1. \tag{12}$$

3.4. Model likelihood

For the hypothesis where process A influences process B, we form the following $2 \times M_t$ matrix $\mathbf{K}_B^t$ at time t:

$$\mathbf{K}_B^t = \begin{bmatrix} k_B^1 & k_B^2 & \dots & k_B^{M_t} \\ I_B^1 & I_B^2 & \dots & I_B^{M_t} \end{bmatrix}, \tag{13}$$

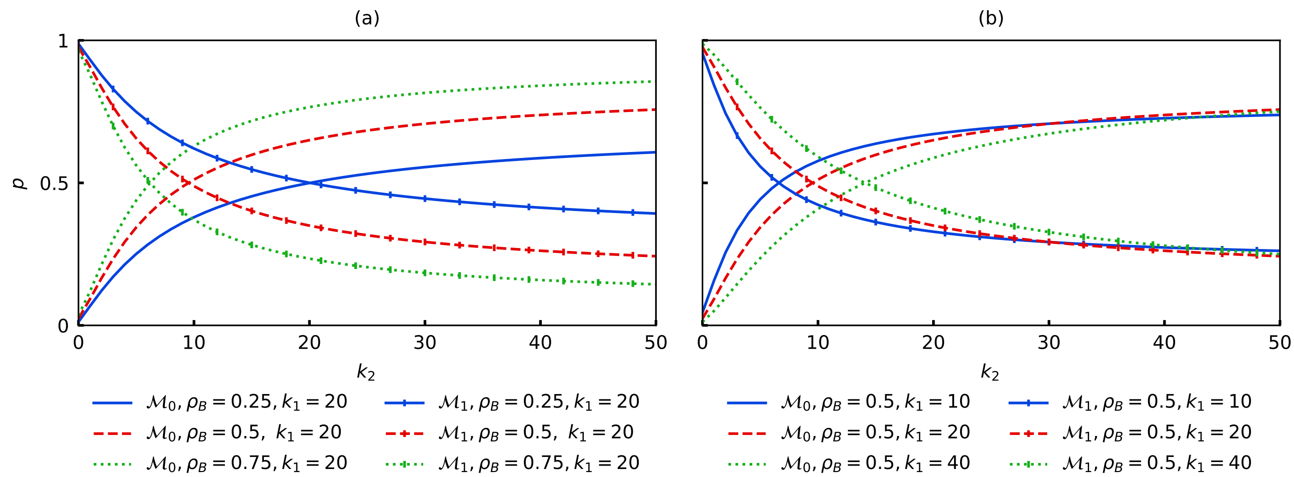

Fig. 2. The likelihoods of $\mathcal{M}_0$ and $\mathcal{M}_1$ for various values of ρ_B, k_1, and k_2.

where k_B^i, $(1 \le i \le M_t)$ is the length of the ith regime for process B, the indicator function $I_B^i = 1$ denotes a regime change due to process A, and M_t is the number of regimes by time t.

While the probability for an arbitrary regime indexed by i when $I_B^i = 0$ can be obtained simply by the YS distribution (1), the probability for regime j given that $I_B^i = 1$ can, by some calculation, be formulated as a Beta function as

$$p(k; \rho, I = 1) = \rho B(\rho, k) \triangleq f'(k; \rho). \tag{14}$$

With (1) and (14), the likelihood can be written as

$$\begin{aligned} p(O_B^{1:t} \mid O_A^{1:t}, \mathcal{M}_i) &= p(K_B^t \mid \mathcal{M}_i) \\ &= \prod_{i=1}^{K} [f(k_B^i; \rho_B)]^{(1-I_B^i)} [f'(k_B^i; \rho_B)]^{I_B^i}, \end{aligned} \tag{15}$$

where for model $\mathcal{M}_0$, $I_B^i = 0$ for all $1 \le i \le M_t$.

To gain further insight about the behavior of the likelihoods of $\mathcal{M}_0$ and $\mathcal{M}_1$, we plot them for various parameters. They are shown in Fig. 2, where the processes A and B have both started new regimes at the beginning of the period. We consider various cases where the process A starts a new regime after k_1 time slots. The symbol k_2 represents the number of time slots when the process B starts a new regime. In plot (a), we show the normalized likelihoods of $\mathcal{M}_0$ and $\mathcal{M}_1$ with fixed k_1 and three different values of ρ_B. In plot (b), we display the same likelihoods for a fixed ρ_B and three different values of k_1.

3.5. Model selection framework with Gibbs sampling

For estimation of ρ_A and ρ_B, we use the Gibbs sampling framework outlined in [13] and adapt it to the interactive case. The Gibbs sampling process for estimating ρ_B can be summarized as follows:

- For $i = 1, \ldots, k$,
  - if $I_B^i = 0$, sample $t_i \mid \rho_B, k_i \sim$ Beta$(\rho_B + 1, k_i)$,
  - if $I_B^i = 1$, sample $t_i \mid \rho_B, k_i \sim$ Beta(ρ_B, k_i);
- Compute $w_i = \log t_i$, for $i = 1, \ldots, K$;
- Sample $\rho \mid \mathbf{w}, \mathbf{k} \sim$ Gamma$(a + n, b + \sum_{i=1}^{n} w_i)$.

where $\mathbf{w}$ and $\mathbf{k}$ denote the sets of all w_i and k_i, respectively. We use Beta(u, v) to denote a Beta distribution with shape parameters u and v, and Gamma(k, θ), a Gamma distribution with shape parameter k and scale parameter θ.

4. A STOCHASTIC MODEL OF INTERACTIONS

Here we introduce a model that allows for a varying strength of influence between processes. According to this model, when an influencing process switches a regime, there is a probability π that the influenced process will reset its counter in the following time instant. We denote this model by $\mathcal{M}_s$.

4.1. Scheme for estimation

In model $\mathcal{M}_s$, in addition to the parameters ρ_A and ρ_B, we have to estimate the probability of influence π. When the influencing process switches its regime, we need to estimate if the influenced process has reset its counter or not. Consider that in the observations there are a total of n time slots when the influencing process has switched its regime. Then there exist 2^n possible traces of the influenced process resetting its counter. We denote these traces by Q_i, where $1 \le i \le 2^n$.

We carry out Gibbs sampling for each given trace and obtain the estimated values of ρ_B. For convenience in notation, we denote the estimated value of ρ_B for trace i as $\rho_{B,i}$. Further, as a prior distribution for π, we assume a Beta distribution given by Beta(α_0, β_0). The posterior distribution of π

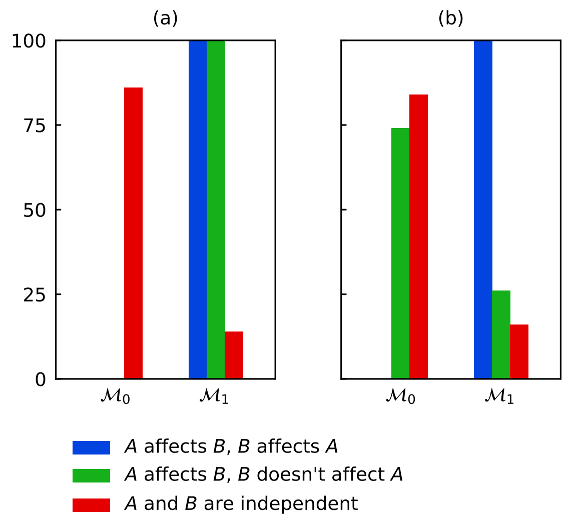

Fig. 3. Simulation results for the $\mathcal{M}_d$ model. The color blue refers to the case when processes A and B affect each other, the green when the process A affects B but B does not affect A, and red when A and B are independent. Plot (a) shows the model selection results when A influences B, and plot (b) when B influences A.

is also a Beta distribution given by $p_i(\pi) = \text{Beta}(\alpha_i, \beta_i)$. The parameters α_i and β_i can be obtained as follows:

$$\alpha_i = \alpha_0 + m_i, \quad (16)$$

$$\beta_i = \beta_0 + n - m_i, \quad (17)$$

where m_i is the number of resets in trace Q_i. For each trace Q_i, we also calculate the likelihood w_i. Given the observations O_A and O_B, and the trace Q_i, we formulate the matrix $\mathbf{K}$ of the influenced process and use (15) for the likelihood of the considered trace.

With this information, we construct a posteriori distribution for ρ_B and π. The a posteriori probability distribution of ρ_B is

$$p(\rho_B) = \sum_{i=1}^{2^n} w_i \delta(\rho - \rho_{B_i}), \quad (18)$$

where $\delta(\cdot)$ is the Dirac delta function located at ρ_{B_i}, and the a posteriori probability distribution of π is given by

$$p(\pi) = \sum_{i=1}^{2^n} w_i \text{Beta}(\alpha_i, \beta_i). \quad (19)$$

5. SIMULATION RESULTS

We conducted experiments to test the proposed methods for inference of the $\mathcal{M}_d$ and $\mathcal{M}_s$ models. In Figure 3, we show the

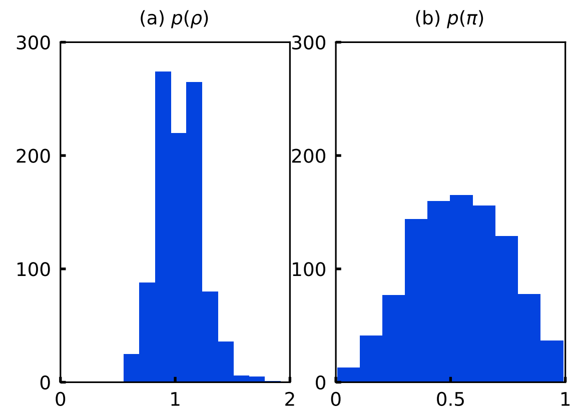

Fig. 4. The simulation results for the $\mathcal{M}_s$ model. Plot (a) shows the a posteriori distribution for ρ, and plot (b) shows the a posteriori distribution for p. The abscissa shows the value of the variables and the ordinate shows the frequency.

Monte Carlo results for $\mathcal{M}_d$. We considered three scenarios of influence between A and B. In the first scenario A and B influenced each other mutually; in the second scenario A affects B but B does not affect A; in the third scenario A and B are independent. For each scenario, we conducted model selection experiments 100 times. The lengths of the time series were 2000 time slots. We ran the Gibbs sampling method with 1000 iterations. The results are shown in Fig. 3. We can see from them that the model selection produces accurate results. With these simulations, we also showed that our model estimation scheme can correctly choose the model when there is a mutual interaction between the processes.

In Fig. 4, we implement the proposed scheme for the $\mathcal{M}_s$ model, and show Monte Carlo results for the a posteriori distributions of ρ_B and π. The true value of ρ_B was 0.75 and the true value of π was 0.5. We can see that the distributions center around the true values of π, but the center of the resulting distribution of ρ_B is somewhat biased and larger than the true value. The length of the time series was 100 samples, and the number of Gibbs sampling iterations was equal to 400.

6. CONCLUSION

In this paper we propose a deterministic and stochastic interactive Yule-Simon models that describe the nature of their interactions. For the deterministic model, we propose a Gibbs sampling-based scheme for estimating the unknown parameters of the Yule-Simon processes and a method for selecting the model that describes the interaction between the processes. For the stochastic model, we propose a trace-based Monte Carlo method to obtain the a posteriori probability distribution of the estimated parameters. Simulation results that demonstrate the performance of the proposed methods are also provided.

7. REFERENCES

[1] M. E. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[2] M. V. Simkin and V. P. Roychowdhury, "Re-inventing Willis," *Physics Reports*, vol. 502, no. 1, pp. 1–35, 2011.

[3] D. Marković and C. Gros, "Power laws and self-organized criticality in theory and nature," *Physics Reports*, vol. 536, no. 2, pp. 41–74, 2014.

[4] J. Saramäki and E. Moro, "From seconds to months: An overview of multi-scale dynamics of mobile telephone calls," *The European Physical Journal B*, vol. 88, no. 6, pp. 1–10, 2015.

[5] E. L. Geist and T. Parsons, "Distribution of tsunami interevent times," *Geophysical Research Letters*, vol. 35, no. 2, 2008.

[6] M. Jagielski, R. Kutner, and D. Sornette, "Theory of earthquakes interevent times applied to financial markets," *Physica A: Statistical Mechanics and its Applications*, vol. 483, pp. 68–73, 2017.

[7] G. U. Yule *et al.*, "A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS," *Phil. Trans. R. Soc. Lond. B*, vol. 213, no. 402-410, pp. 21–87, 1925.

[8] H. A. Simon, "On a class of skew distribution functions," *Biometrika*, vol. 42, no. 3/4, pp. 425–440, 1955.

[9] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[10] D. d. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, vol. 27, no. 5, pp. 292–306, 1976.

[11] R. K. Merton, "The Matthew effect in science: The reward and communication systems of science are considered," *Science*, vol. 159, no. 3810, pp. 56–63, 1968.

[12] A. A. Hensley and P. M. Djurić, "Nonparametric learning for hidden Markov models with preferential attachment dynamics," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3854–3858.

[13] F. Leisen, L. Rossini, and C. Villa, "A note on the posterior inference for the Yule–Simon distribution," *Journal of Statistical Computation and Simulation*, vol. 87, no. 6, pp. 1179–1188, 2017.