Author: Ksenia N. Kyzyurova

Stochastic processes

October 20, 2019

This monograph is the most concise and complete introduction to stochastic processes.

Definition. A stochastic process ξ_υ is defined in the literature as a collection of infinite number of random variables $\{\xi_\upsilon\}$ indexed by $\upsilon \in X$, a countable or an uncountable set, whose conditional and joint and therefore, marginal distributions of any subset of finite number of these variables are valid and compatible, that is, their corresponding cumulative probability density functions $F_{\{\cdot\}}$ must satisfy the following conditions

$$\begin{aligned} F_{\xi_1,\xi_2,\ldots,\xi_{k+1}}(\xi_1 < z_1, \ldots, \xi_k < z_k, \xi_{k+1} < \infty) = F_{\xi_1,\xi_2,\ldots,\xi_k}(\xi_1 < z_1, \ldots, \xi_k < z_k)\,, \\ F_{\xi_{s_1},\xi_{s_3}|\xi_{s_2}}(\xi_{s_1} < z_{s_1}, \xi_{s_3} < \infty \mid \xi_{s_2}) = F_{\xi_{s_1}|\xi_{s_2}}(\xi_{s_1} < z_{s_1} \mid \xi_{s_2})\,, \\ F_{\xi_{s_1},\xi_{s_2}}(\xi_{s_1} < z_{s_1}, \xi_{s_2} < z_{s_2}) = F_{\xi_{s_1}|\xi_{s_2}}(\xi_{s_1} < z_{s_1} \mid \xi_{s_2} < z_{s_2}) F_{\xi_{s_2}}(\xi_{s_2} < z_{s_2})\,, \end{aligned} \tag{1}$$

where s_1, s_2 and s_3 are disjoint subsets of random variables from the collection $\{\xi_\upsilon\}$. The process ξ_υ may also be equivalently denoted as $\xi(\upsilon)$. (υ is the Greek letter Upsilon.)

The word "process" means that there is the development of events which are not independent from each other this way or another. Therefore, the collection of dependent random variables, which satisfy (1) is called the process. The infinite collection of random variables which depend only on index υ, but are independent pairwise, is also called a stochastic process. A collection of independent random variables which do not depend on index υ either, satisfies the definition of a stochastic process, although the essential idea of the "process" is lost.

If $\upsilon \in X \subset \mathbb{R}$ comes from a subset of real numbers, e.g. $X = (-\infty, \infty)$ or $X = [0, \infty)$, we call ξ_υ a stochastic process in continuous time. One example of such a process is a Gaussian process. This process is well-introduced in the monograph [2] and research [3,4], where other constructions based on Gaussian processes, a censored Gaussian Process and a T-process are given. In this monograph we consider stochastic processes indexed by υ from a *countable* set. The simplest construction uses $X = \mathbb{N}_{\geq 0} = \{0, 1, 2, \ldots\}$, a set of natural numbers. The mathematical model of division of cells is considered and serves as an example of such processes.

A collection of random variables $\zeta_\iota, \iota \in X = \mathbb{N}_{\geq 0}$ is called a stochastic process in discrete time, or a *discrete time series* with equidistant lags which are all equal to one. One example of such series is considered in [1]. Another example is provided in this monograph.

Copyright © by Ksenia N. Kyzyurova
All rights reserved

1 Mathematical model of division of cells

The model of the division of cells which is developed in this monograph is a convenient mathematical construction for illustration of the notion of a stochastic process.

Deterministic model. Suppose that a cell divides with some constant rate, say, providing u cells per unit time. That is, e.g., either a cell provides u new cells and disappears itself at each unit time; or a cell provides $u-1$ new cells and stays alive, therefore, having the total amount of generated cells equal to u cells per unit time.

Proposition. The number of cells at t units of time is $u^t = 2^{(\log_2 u)t}$, resulting in exponential growth of the total number of cells. *Proof.* Relies on combinatorics. ■

This model may be considered as a simple mathematical model which deterministically finds the total number of cells at time t given the rate of division, the number of cells produced by a single cell per unit time. The unit time is relative: u cells per unit time is equivalent to, for example, 2 cells per $\log_2 u$ units of time or ℓ cells per $\log_\ell u$ units of time.

Probabilistic modeling. In nature we observe variation in outcomes of the same phenomenon which more or less are developed in the same conditions. The idea is to account for this variation by introducing probabilistic modeling of the division of cells, in place of the deterministic approach. The attempt is to construct a simple probabilistic model of the division of cells.

Let the process of division start with one cell, that is at time zero $\zeta_0 = 1$ cell. Let each cell divide with probability $0 < c < 1$, stay the same with probability $0 < b < 1$ and die with probability $a = 1 - b - c$. Formally, the number of cells at time point $\iota = 1$ is a random variable ζ_1

$$\zeta_1 = \begin{cases} 0\,, & a\,, \\ 1\,, & b\,, \\ 2\,, & c\,, \end{cases} \tag{2}$$

where $(a, b, c)^{\mathrm{T}}$ is a simplex vector of probabilities of order two. Then, the distribution of the number of cells at any time point may be deduced.

Lemma. The conditional distribution of the number of cells ζ_ι at any other index $\iota \geq 1$, given the number of cells at the previous time step $\iota - 1$, is

$$P(\zeta_\iota \mid \zeta_{\iota-1}) = \begin{cases} \sum_{j=0}^{(\zeta_\iota - \zeta_\iota \bmod 2)/2} C_{\zeta_{\iota-1}}^{\zeta_{\iota-1}-\zeta_\iota+j} a^{\zeta_{\iota-1}-\zeta_\iota+j} C_{\zeta_\iota - j}^{\zeta_\iota - 2j} b^{\zeta_\iota - 2j} (1-a-b)^j\,, & \zeta_\iota \leq \zeta_{\iota-1}\,; \\ \sum_{j=0}^{(\zeta'_\iota - \zeta'_\iota \bmod 2)/2} C_{\zeta_{\iota-1}}^{\zeta_{\iota-1}-\zeta'_\iota+j} (1-a-b)^{\zeta_{\iota-1}-\zeta'_\iota+j} C_{\zeta'_\iota - j}^{\zeta'_\iota - 2j} b^{\zeta'_\iota - 2j} a^j\,, & \zeta_\iota > \zeta_{\iota-1}\,, \end{cases} \tag{3}$$

where $\zeta_\iota \in \Omega_{\iota|\iota-1} = \{0, \ldots, 2\zeta_{\iota-1}\}$, the conditional space of outcomes at time point ι; $j \in \mathbb{N}_{\geq 1}$; $\zeta'_\iota = 2\zeta_{\iota-1} - \zeta_\iota$. [1] [2] [3] *Proof.* Relies on combinatorics. ■

Theorem. $\zeta(\iota) = \zeta_\iota, \iota \in X = \mathbb{N}_{\geq 0}$ is a stochastic process in discrete time. *Proof.* ζ_ι satisfies (1) by construction. ■

Theorem. The conditional distribution (3) is stationary. *Proof.* The expression of the conditional distribution does not depend on ι, but only on the values of ζ_ι and $\zeta_{\iota-1}$ themselves. ■

Theorem. If $a \neq c$, then the process ζ_ι is non-stationary in any sense. *Proof.* Distributional characteristics (such as, cumulative distribution function, expectation, correlation among pairs of variables) of the process ζ_ι change with ι; that is, they do not stay the same for any positive time lag λ. *Proof.* Consider the distributional characteristics to see that this is indeed so. ■

Theorem. If $a = c$, then the process is stationary in the mean, but not in correlation or cumulative distribution function. *Proof.* Relies on results in section 7.1 which discusses this special case. ■

Theorem. The process ζ_ι is a discrete time Markov chain. *Proof.* The probability of the number of cells conditionally depends only on the number of cells observed at the previous time point; and does not depend on the numbers of cells observed at any other time point. ■

Lemma. The marginal probabilities for all $\iota \geq 1$

$$P(\zeta_\iota = \gamma) = \sum_{j=(\gamma + \gamma \bmod 2)/2}^{2^{\iota-1}} P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = j) P(\zeta_{\iota-1} = j)\,, \tag{4}$$

where $\zeta_\iota \in \Omega_\iota = \{0, 1, \ldots, 2^\iota\}$, the marginal space of outcomes at time ι. *Proof.* Since $P\left(\zeta_{\iota-1} = j < \frac{\gamma + \gamma \bmod 2}{2} \mid \zeta_\iota = \gamma\right) = 0$, the lemma holds true. ■

[1] $C_w^y = \frac{y!}{w!(w-y)!}$ is the number of unordered selections of y objects out of w objects.
[2] $m \bmod u$ is the remainder from the division of $m \in \mathbb{N}_{\geq 0}$ over $u \in \mathbb{N}_{\geq 1}$.
[3] The expression (3) implicitly assumes that $0^0 = 1$.

2 The model of division of cells with the restriction on the total amount of cells

Suppose we are interested in developing a model such that at any given time point of the division process there exists a limit on the total amount of cells possible to occur. Such a model is convenient to construct building on the results from the model in the previous section. In order to formulate the model and describe the corresponding stochastic process, we must find its conditional probability of the number of cells ζ_ι^ψ occurring at time point ι given the number of cells occurring at time point $\iota - 1$, while incorporating the information about the limit $\max \zeta_\iota = \psi$ for all $\iota \in \mathbb{N}_{\geq 0}$, where $\psi \in \mathbb{N}_{\geq 1}$.

Lemma. In the presence of the limit ψ on the total amount of cells occurring at any time point ι and denoted as ζ_ι^ψ, the following conditional distribution occurs

$$P(\zeta_\iota^\psi = \gamma \mid \zeta_{\iota-1}^\psi = \varkappa) = \begin{cases} P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa \leq \psi, \max \zeta_\iota = \psi) = \frac{P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa)}{\sum_{j=0}^{\min(\psi, 2^\iota, 2\varkappa)} P(\zeta_\iota = j \mid \zeta_{\iota-1} = \varkappa)}\,, \\ P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa > \psi, \max \zeta_\iota = \psi) = 0\,, \end{cases} \tag{5}$$

where for each $\iota \in \mathbb{N}_{\geq 1}$ the number of cells $\zeta_\iota^\psi \in \Omega_{\iota|\iota-1,\psi} = \{0, 1, \ldots, \min(\psi, 2^\iota, 2\zeta_{\iota-1})\}$ is the conditional space of outcomes in the presence of the limit ψ. *Proof.* Since $P(\zeta_\iota = j > \min(\psi, 2^\iota, 2\varkappa) \mid \zeta_{\iota-1} = \varkappa) = 0$, the lemma holds true. ■

Algorithm C1 in section 8 prototypes the programmed solution for computation of the conditional probabilities.

Lemma. The marginal probabilities of the number of cells ζ_ι^ψ to occur in the presence of the limit ψ at time point ι are

$$P(\zeta_\iota^\psi = \gamma) = \begin{cases} P(\zeta_\iota = \gamma \mid \max \zeta_\iota = \psi < 2^\iota, (\gamma + \gamma \bmod 2)/2 \leq \psi) = \\ \quad = \displaystyle\sum_{\varkappa = \frac{\gamma + \gamma \bmod 2}{2}}^{\min(2^{\iota-1}, \psi)} P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa) P(\zeta_{\iota-1} = \varkappa \mid \max \zeta_{\iota-1} = \psi)\,, \\ P(\zeta_\iota = \gamma \mid \max \zeta_\iota = \psi \geq 2^\iota) = P(\zeta_\iota = \gamma)\,, \\ P(\zeta_\iota = \gamma \mid \max \zeta_\iota = \psi < 2^\iota, (\gamma + \gamma \bmod 2)/2 > \psi) = 0\,, \end{cases} \tag{6}$$

where for each $\iota \in \mathbb{N}_{\geq 1}$ the number of cells $\zeta_\iota^\psi \in \Omega_{\iota,\psi} = \{0, 1, \ldots, \min(\psi, 2^\iota)\}$ is the marginal space of outcomes in the presence of the limit ψ. *Proof.* By construction. ■

Algorithm C2 in section 8 provides the programmed solution for computation of the marginal probabilities.

Theorem. $\zeta^\psi(\iota) = \zeta_\iota^\psi, \iota \in X = \mathbb{N}_{\geq 0}$ is a stochastic process in discrete time. *Proof.* ζ_ι^ψ satisfies (1) by construction. ■

3 Computational lemmas

For the implementation of the models (with and without a restriction ψ) the following computational lemmas have been found to be useful for enabling computations and/or speeding up the computing time of the conditional and marginal probabilities of interest.

Lemma. If $\zeta_{\iota-1} - \zeta_\iota + j > 0$, then

$$P(\zeta_\iota \mid \zeta_{\iota-1}) = a(1 + (\zeta_\iota - j)/(\zeta_{\iota-1} - \zeta_\iota + j)) P(\zeta_\iota \mid \zeta_{\iota-1} - 1)\,. \tag{7}$$

Lemma. Expressions of the type $C_m^{m-\zeta} a^{m-\zeta} b^\zeta$ are simplified for computations as

$$C_m^{m-\zeta} a^{m-\zeta} b^\zeta = \begin{cases} a^{m-2\zeta} \displaystyle\prod_{j=1}^{\zeta} \left(ab \left(1 + \frac{m-\zeta}{j} \right) \right), & m > 2\zeta\,; \\ (1-a)^{2\zeta} \displaystyle\prod_{j=1}^{m-\zeta} \left(ab \left(1 + \frac{\zeta}{j} \right) \right), & m \leq 2\zeta\,. \end{cases} \tag{8}$$

Lemma. Computational simplifications of the extreme cases, such that the simplex vector of initial probabilities $(a, b, c = 1 - a - b)^{\mathrm{T}}$ of order 2 is reduced to the vector of order 1

$$\begin{aligned} P(\zeta_\iota \mid \zeta_{\iota-1}, 1 - a - b = 0) &= \begin{cases} C_{\zeta_{\iota-1}}^{\zeta_{\iota-1} - \zeta_\iota} a^{\zeta_{\iota-1} - \zeta_\iota} b^{\zeta_\iota}\,, & \text{if } \zeta_{\iota-1} \geq \zeta_\iota\,; \\ 0\,, & \text{otherwise.} \end{cases} \\ P(\zeta_\iota \mid \zeta_{\iota-1}, a = 0) &= \begin{cases} C_{\zeta_{\iota-1}}^{2\zeta_{\iota-1} - \zeta_\iota} b^{2\zeta_{\iota-1} - \zeta_\iota} (1-b)^{\zeta_\iota - \zeta_{\iota-1}}\,, & \text{if } \zeta_{\iota-1} \leq \zeta_\iota \leq 2\zeta_{\iota-1}\,; \\ 0\,, & \text{otherwise.} \end{cases} \\ P(\zeta_\iota \mid \zeta_{\iota-1}, b = 0) &= \begin{cases} C_{\zeta_{\iota-1}}^{\zeta_{\iota-1} - \zeta_\iota/2} a^{\zeta_{\iota-1} - \zeta_\iota/2} (1-a)^{\zeta_\iota/2}\,, & \text{if } \zeta_\iota \leq 2\zeta_{\iota-1} \text{ and } \zeta_\iota \bmod 2 = 0\,; \\ 0\,, & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

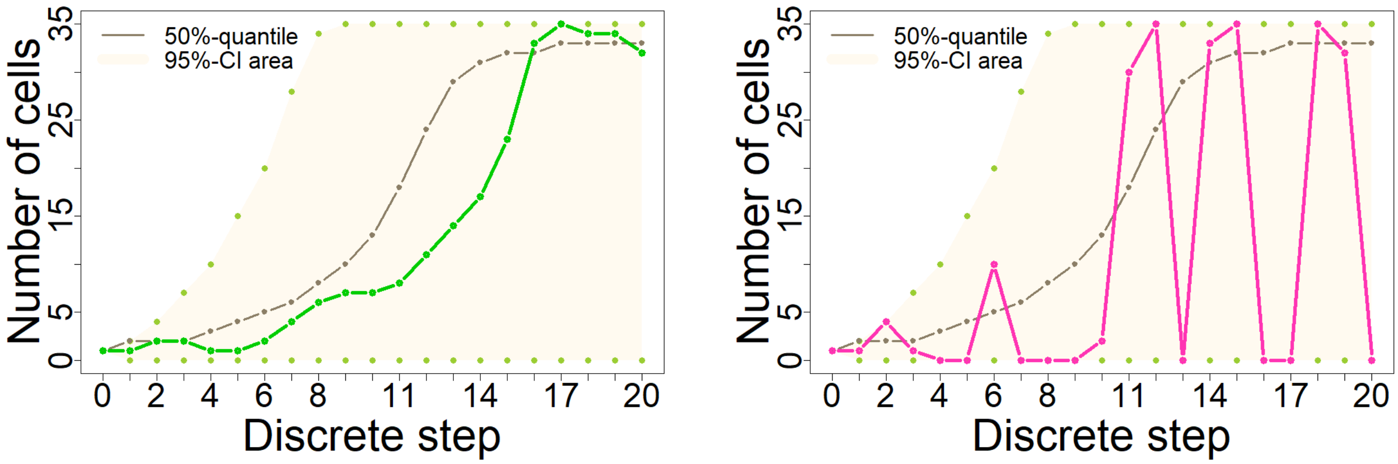

Figure 1: Left plot: bright green connected points is a random sample generated from the model of division of cells specified by the simplex vector of probabilities with $a = 0.18750$ and $b = 0.28125$, the limit $\psi = 36$. The observed time period $T = 20$. Light green pairs of points are 2.5% and 97.5% quantiles of the distribution of the number of cells at each time point $\iota \in \{0, 1, \ldots, T\}$. Light brown (beige) area is the 95% central credible area enclosed between the selected quantiles. The dark brown connected points are the median of the distribution of the number of cells. Right plot: bright pink connected points is a sample from the marginal distributions of the same model over index $\iota \in \{0, 1, \ldots, T\}$. The rest of the plot coincides with the left plot.

Lemma. The conditional probabilities depending on various values of the limit ψ and the number of cells at the previous step $\zeta^{\psi}_{\iota-1}$ are simplified

$$P(\zeta^{\psi}_{\iota} = \gamma \mid \zeta^{\psi}_{\iota-1} = \varkappa) = \begin{cases} P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa \le \psi, \psi \ge 2^{\iota}) = P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa)\,, \\ P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa, \psi < 2^{\iota}, 2\varkappa \le \psi) = P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa)\,, \\ P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa \le \psi, \psi < 2^{\iota}, 2\varkappa > \psi) = \frac{P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa)}{\sum_{j=0}^{\psi} P(\zeta_\iota = \alpha \mid \zeta_{\iota-1} = \varkappa)}\,. \end{cases} \tag{10}$$

4 Generation of a random sample path from the models

One possibility to obtain an example of how the number of cells are evolved in time, a random sample from the model, is to simply generate a random variable which gives zero, one or two cells at time ι (with fixed probabilities a, b, c) for every cell at time $\iota - 1$ and calculate the sum of the outcomes to get the total number of cells. If the total number of cells at time ι occurs to be greater than the limit ψ, this means, that at time ι an impossible event (the number of cells greater than the limit) has occurred. Thus, one must discard this sample and repeat the generation process until a possible event (the number of cells less than or equal to the limit) occurs. The necessity to discard samples slows down the computations. In order to avoid this burden, one relies on the derived formulae of conditional probabilities (3) and (5) and computational lemmas (7), (8), (9), (10) for the implementation of the model. Doing so, one obtains a true *sample path* $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_T)$ from the model with the restriction ψ, where $\omega_\iota \in \Omega_{\iota|\iota-1,\psi}$ for $\iota \in \{1, 2, \ldots, T\}$. T is the observed time index period.

An example of a random sample from the model of division of cells with the fixed probability vector $(a = 0.18750, b = 0.28125, c = 0.53125)^{\mathrm{T}}$, the limit $\psi = 36$ and the time period $T = 20$ is shown in Figure 1, left plot.

Producing a sample by using marginal distributions only, that is $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_T)$, such that $\nu_\iota \in \Omega_{\iota,\psi}$; that is ν_ι is independent from $\nu_{\iota' \neq \iota}$, one neglects the correlation structure among variables, and the sample $\boldsymbol{\nu}$ is not from the model. The representation of the sample path is its *trajectory*, which is lost if the variables are independent over index $\iota \in X = \{0, 1, \ldots\}$. An example of a sample $\boldsymbol{\nu}$ is given on the right plot of Figure 1. Analogous loss of correlation structure for the processes in continuous time is discussed and illustrated in [5].

5 Model of division of cells with other rates of division

Suppose that a model considered in section 1 now has the probability rate of division of cells such that a single cell produces up to $u \in \mathrm{N}_{\ge 3}$ number of cells per unit time.

Theorem. Let the number of produced cells by a single cell in one time step be a random variable whose outcomes $\alpha = \{0, 1, \ldots, u\}$ are distributed with the simplex probability vector $a_{0:u} = \{a_0, a_1, \ldots, a_u\}^{\intercal}$, then the conditional probability of the number of cells

$$P(\zeta_\iota^u \mid \zeta_{\iota-1}^u) = \begin{cases} \displaystyle\sum_{\boldsymbol{\tau}:(\boldsymbol{\tau},\boldsymbol{\alpha})=\zeta_\iota^u} C_{\zeta_{\iota-1}}^{\tau_0} a_0^{\tau_0} C_{\zeta_{\iota-1}-\tau_0}^{\tau_1} a_1^{\tau_1} \cdot \ldots \cdot C_{\zeta_{\iota-1}-\sum_{j=0}^{u-2}\tau_j}^{\tau_{u-1}} a_{u-1}^{\tau_{u-1}} a_u^{\tau_u}\,, & 0 \le \zeta_\iota^u \le u\zeta_{\iota-1}^u\,; \\ 0\,, & \text{otherwise,} \end{cases} \tag{11}$$

where $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_u)$ with $\tau_j \in \mathbb{N}_{\ge 0}$ for all j and are subject to the constraint $\sum_{j=0}^{u} \tau_j = \zeta_{\iota-1}^u$. *Proof.* The formula is derived using combinatorics. ■

Theorem. The marginal probability of the number of cells ζ_ι^u to occur at time ι is

$$P(\zeta_\iota^u = \gamma) = \sum_{j=\frac{\gamma+r}{u}}^{u^{\iota-1}} P(\zeta_\iota^u = \gamma \mid \zeta_{\iota-1}^u = j) P(\zeta_{\iota-1}^u = j) \tag{12}$$

for all $\iota \ge 1$ and $u \ge 2$, where $r \in \{0, 1, \ldots, u-1\}$ is such that $(\gamma + r) \bmod u = 0$. *Proof.* $P\left(\zeta_{\iota-1}^u = j < \frac{\gamma+r}{u} \mid \zeta_\iota^u = \gamma\right) = 0$. ■

Theorem. $\zeta^u(\iota) = \zeta_\iota^u, \iota \in X = \mathbb{N}_{\ge 0}$ is a stochastic process in discrete time. *Proof.* ζ_ι^u satisfies (1) by construction. ■

5.1 Algorithm for identifying vectors τ in the expression (11)

Denote as $\Upsilon_{\iota|\iota-1}^u$ the desired set of all vectors $\boldsymbol{\tau}$ with non-negative integer-valued entries that satisfy $(\boldsymbol{\tau}, \boldsymbol{\alpha}) = \zeta_\iota^u$ and $\sum_{j=0}^{u} = \zeta_{\iota-1}^u$. Define the following sets of elements

$$\begin{gathered} \Xi_u = \left\{0, 1, \ldots, \min\left\{\zeta_{\iota-1}^u, \lfloor \zeta_\iota^u / u \rfloor\right\}\right\}\,, \\ \Xi_{u-k+1} = \left\{0, 1, \ldots, \min\left\{\zeta_{\iota-1}^u - \sum_{j=u-k+2}^{u} \tau_j, \lfloor(\zeta_\iota^u - \sum_{j=u-k+2}^{u} j\tau_j)/u\rfloor\right\}\right\} \end{gathered} \tag{13}$$

for all $k = 2, 3, \ldots, u-1$.[4]

Algorithm

(1) Start with $k = 1$.

(2) For each element $\tau_{u-k+1} \in \Xi_{u-k+1}$

if $\zeta_\iota^u - \sum_{j=u-k+1}^{u} j\tau_j \ge 0$ and $\zeta_{\iota-1}^u - \sum_{j=u-k+1}^{u} \tau_j \ge 0$ are both true

then there may exist a vector $\boldsymbol{\tau}$ *whose entries* $\tau_{u-k+1}, \ldots, \tau_u$ *form part of a vector from the set* $\Upsilon_{\iota|\iota-1}^u$

if $u - k + 1 > 2$

store in memory the current value of τ_{u-k+1}

increase the value of k by one, that is $k := k + 1$

repeat the procedure (2) for searching the next entry of a vector $\boldsymbol{\tau}$

else {

$\tau_1 = \zeta_\iota^u - \sum_{j=2}^{u} j\tau_j$

if $\tau_1 \ge 0$ then $\tau_0 = \zeta_{\iota-1}^u - \sum_{j=1}^{u} \tau_j$

if $\tau_0 \ge 0$ then we have found a valid sample $\boldsymbol{\tau} = \{\tau_0, \tau_1, \ldots, \tau_u\}$ from the desired set $\Upsilon_{\iota|\iota-1}^u$.

}

end

[4] $\lfloor m/u \rfloor = \frac{m - m \bmod u}{u}$.

6 Model in section 5 with the limit on the total amount of cells

Lemma. The following conditional distribution is a generalization to the number of cells u occurring with a vector of probabilities

$$P(\zeta_\iota^{u,\psi}=\gamma \mid \zeta_{\iota-1}^{u,\psi}=\varkappa)=P(\zeta_\iota=\gamma \mid \zeta_{\iota-1}=\varkappa\leq\psi,\psi<\infty)=\begin{cases}\frac{P(\zeta_\iota=\gamma|\zeta_{\iota-1}=\varkappa)}{\sum_{j=0}^{\min(\psi,u^\iota,u\varkappa)}P(\zeta_\iota=j|\zeta_{\iota-1}=\varkappa)}\,, & \psi\leq u\varkappa\,;\\ P(\zeta_\iota=\gamma \mid \zeta_{\iota-1}=\varkappa)\,, & \psi>u\varkappa\,.\end{cases} \tag{14}$$

Theorem. The marginal probability of the number of cells $\zeta_\iota^{u,\psi}$ to occur at time ι is

$$P(\zeta_\iota^{u,\psi}=\gamma)=\begin{cases}P(\zeta_\iota=\gamma \mid \max\,\zeta_\iota=\psi,(\gamma+r)/u\leq\psi)=\\ \quad=\displaystyle\sum_{j=\frac{\gamma+r}{u}}^{\min(u^{\iota-1},\psi)}P(\zeta_\iota=\gamma \mid \zeta_{\iota-1}=j,\max\,\zeta_\iota=\psi)P(\zeta_{\iota-1}=j \mid \max\,\zeta_{\iota-1}=\psi)\,,\\ P(\zeta_\iota=\gamma \mid \max\,\zeta_\iota=\psi\geq u^\iota)=P(\zeta_\iota=\gamma)\,,\\ P(\zeta_\iota=\gamma \mid \max\zeta_\iota=\psi<u^\iota,(\gamma+r)/u>\psi)=0\,.\end{cases} \tag{15}$$

Theorem. $\zeta^{u,\psi}(\iota)=\zeta_\iota^{u,\psi},\iota\in X=\mathbb{N}_{\geq 0}$ is a stochastic process in discrete time. *Proof.* $\zeta_\iota^{u,\psi}$ satisfies (1) by construction. ■

Theorem. Let $w>u$, then $\zeta_\iota^{u,\psi}\to\zeta_\iota^{w,\psi},\iota\in X=\mathbb{N}_{\geq 0}$ in distribution as $a_{(w+1):u}\to\mathbf{0}$. *Proof.* By construction. ■

Several analogous theorems as the ones regarding the stochastic process $\zeta_\iota^{u=2,\psi=\infty}=\zeta_\iota^{\psi=\infty}=\zeta_\iota^{u=2}=\zeta_\iota$ are generalizable to the processes $\zeta_\iota^{\psi},\zeta_\iota^{u}$ and $\zeta_\iota^{u,\psi}$.

Lemma. **(a)** In the presence of two types of cells and assuming each type of cell is independent from another

$$P(\zeta_\iota,\eta_\iota \mid \zeta_{\iota-1},\eta_{\iota-1},\psi)=\frac{P(\zeta_\iota \mid \zeta_{\iota-1})P(\eta_\iota \mid \eta_{\iota-1})}{\sum_{\zeta_\iota+\eta_\iota=0}^{\psi}P(\zeta_\iota \mid \zeta_{\iota-1})P(\eta_\iota \mid \eta_{\iota-1})}\,. \tag{16}$$

(b) If the number of different types of cells is $S>2$ with corresponding processes for each type of cells denoted as $\{\phi_\iota^s\}$ for $s\in\{1,2,\dots,S\}$, then

$$P(\phi_\iota^1,\dots,\phi_\iota^S \mid \phi_{\iota-1}^1,\dots,\phi_{\iota-1}^S,\psi)=\frac{\displaystyle\prod_{s=1}^{S}P(\phi_\iota^s \mid \phi_{\iota-1}^s)}{\displaystyle\sum_{\phi_\iota^1+\ldots+\phi_\iota^S=0}^{\psi}\prod_{s=1}^{S}P(\phi_\iota^s \mid \phi_{\iota-1}^s)}\,. \tag{17}$$

7 Appendix

The following is a set of auxiliary mathematical statements which provide details on the considered models. The proofs are omitted.

Lemma. Distributions $P(\zeta_\iota \mid \zeta_{\iota-1})$ given by (3) may be either uni- or multi-modal.

Proposition. Alternative form for the expression (3) is

$$P(\zeta_\iota \mid \zeta_{\iota-1})=\begin{cases}\sum_{j=0}^{(\zeta_\iota-\zeta_\iota \bmod 2)/2}\frac{\zeta_{\iota-1}!a^{\zeta_{\iota-1}-\zeta_\iota+j}b^{\zeta_\iota-2j}(1-a-b)^j}{(\zeta_{\iota-1}-\zeta_\iota+j)!(\zeta_\iota-2j)!j!}\,, & \zeta_\iota\leq\zeta_{\iota-1}\,,\\ \sum_{j=0}^{(\zeta'_\iota-\zeta'_\iota \bmod 2)/2}\frac{\zeta_{\iota-1}!c^{\zeta_{\iota-1}-\zeta'_\iota+j}b^{\zeta'_\iota-2j}a^j}{(\zeta_{\iota-1}-\zeta'_\iota+j)!(\zeta'_\iota-2j)!j!}\,, & \text{otherwise},\end{cases} \tag{18}$$

where $\zeta'_\iota=2\zeta_{\iota-1}-\zeta_\iota$.

Lemma. **(a)** Alternative form for the expression (4) of the marginal probability is

$$P(\zeta_\iota=2^\iota-\beta)=\sum_{j=0}^{(\beta-\beta \bmod 2)/2}P(\zeta_\iota=\beta-2j,\zeta_{\iota-1}=2^{\iota-1}-j)\,, \tag{19}$$

where $P(\zeta_\iota=\beta-2j,\zeta_{\iota-1}=2^{\iota-1}-j)=P(\zeta_\iota=\beta-2j \mid \zeta_{\iota-1}=2^{\iota-1}-j)P(\zeta_{\iota-1}=2^{\iota-1}-j)$.

(b) In particular,

$$\begin{aligned}P(\zeta_\iota=2^\iota)&=c^{\sum_{j=0}^{\iota-1}2^j}\,,\\ P(\zeta_\iota=2^\iota-1)&=2^{\iota-1}bc^{\sum_{j=0}^{\iota-1}2^j-1}\,,\\ P(\zeta_\iota=2^\iota-2)&=\left(2^{\iota-1}ac^{2^{\iota-1}-1}+2^{\iota-2}\left(2^{\iota-1}-1\right)b^2c^{2^{\iota-1}-2}+2^{\iota-2}bc^{2^{\iota-1}-2}\right)c^{\sum_{j=0}^{\iota-2}2^j}\,,\\ P(\zeta_\iota=2^\iota-3)&=\left(2^{\iota-1}\left(2^{\iota-1}-1\right)\left(2^{\iota-1}-2\right)c^{2^{\iota-1}-3}b^3/6+2^{\iota-1}\left(2^{\iota-1}-1\right)c^{2^{\iota-1}-2}ba+\left(2^{\iota-1}-1\right)b^22^{\iota-2}c^{2^{\iota-1}-3}\right)c^{\sum_{j=0}^{\iota-2}2^j}\,.\end{aligned} \tag{20}$$

(c) At time $\iota = 2$

$$\begin{aligned}
&P(\zeta_2 = 0) = a + ab + a^2 c\,,\\
&P(\zeta_2 = 1) = b^2 + 2abc\,,\\
&P(\zeta_2 = 2) = bc + b^2 c + 2ac^2\,,\\
&P(\zeta_2 = 3) = 2bc^2\,,\\
&P(\zeta_2 = 4) = c^3\,.
\end{aligned} \tag{21}$$

Theorem. Expectation, second moment and variance are customarily defined as

$$\begin{aligned}
\mathrm{E}\zeta_\iota &= \sum_{\varkappa=0}^{u^\iota} \varkappa P(\zeta_\iota = \varkappa)\,,\\
\mathrm{E}\zeta_\iota^2 &= \sum_{\varkappa=0}^{u^\iota} \varkappa^2 P(\zeta_\iota = \varkappa)\,,\\
\mathrm{V}\zeta_\iota &= \mathrm{E}\zeta_\iota^2 - (\mathrm{E}\zeta_\iota)^2\,.
\end{aligned} \tag{22}$$

Correlation and autocovariance are defined as

$$\begin{aligned}
\mathrm{Corr}\,(\zeta_{\iota-1}\zeta_\iota) &= \frac{\mathrm{E}\zeta_{\iota-1}\zeta_\iota - \mathrm{E}\zeta_{\iota-1}\mathrm{E}\zeta_\iota}{\sqrt{\mathrm{V}\zeta_{\iota-1}\mathrm{V}\zeta_\iota}}\,,\\
\mathrm{E}\zeta_{\iota-1}\zeta_\iota = \mathrm{E}_{\zeta_{\iota-1}}\zeta_{\iota-1}(\mathrm{E}\zeta_\iota \mid \zeta_{\iota-1}) &= \sum_{\gamma=0}^{u^{\iota-1}} \gamma \left(\sum_{\varkappa=0,\{\varkappa:\frac{\varkappa+\delta}{u}\leq\gamma\}}^{2\gamma} \varkappa P(\zeta_\iota = \varkappa \mid \gamma) \right) P(\zeta_{\iota-1} = \gamma)\,,
\end{aligned} \tag{23}$$

where $\delta : (\varkappa + \delta) \bmod u = 0$.

The conditional expectation is

$$\mathrm{E}\zeta_\iota \mid \zeta_{\iota-1} = \sum_{\gamma=0}^{2\zeta_{\iota-1}} \gamma P(\zeta_\iota = \gamma \mid \zeta_{\iota-1})\,. \tag{24}$$

7.1 Symmetry

In this part the symmetric case, such that probability of a single cell to die is a and probability of a cell to duplicate at the next time point $c = a$ are equal. In other words the birth and death rates coincide.

Lemma. Conditional distributions (3) $P(\zeta_k \mid \zeta_{k-1})$ are not symmetric in general.

Lemma. The initial distribution is symmetric (that is $a = c$) if and only if all conditional distributions given by (25) are symmetric.

$$P(\zeta_\iota = \gamma \mid \zeta_{\iota-1} = \varkappa) = P(\zeta_\iota = 2\varkappa - \gamma \mid \zeta_{\iota-1} = \varkappa) = \sum_{j=0}^{(\gamma-\gamma \bmod 2)/2} C_\varkappa^{\varkappa-\gamma+j} C_{\gamma-j}^{\gamma-2j} a^\varkappa (1/a - 2)^{\gamma-2j} \text{ if } \gamma \leq \varkappa\,. \tag{25}$$

Lemma. If $a = c$,

$$\mathrm{E}\zeta_\iota \mid \zeta_{\iota-1} = \zeta_{\iota-1}\,. \tag{26}$$

Theorem. If $a = c$, for all $\iota \in \{0, 1, \ldots\}$ expectation

$$\mathrm{E}\zeta_\iota = 1\,. \tag{27}$$

is constant over time index ι.

For all $\iota \geq 2$ variance

$$\mathrm{V}\zeta_\iota = \iota \mathrm{V}\zeta_1 \tag{28}$$

is growing linearly with time. Variance at time point $\iota = 1$ is $\mathrm{V}\zeta_1 = 2a$.

Theorem. If $a = c$, several other moments and cumulants are

$$\begin{aligned} \mathrm{E}(\zeta_\iota - \mathrm{E}\zeta_\iota)^3 &= \frac{(\iota+1)(\iota+2)}{2} 12a^2, \text{ for } \iota \geq 3\,, \\ \mathrm{E}(\zeta_2 - \mathrm{E}\zeta_2)^3 &= 12a^2\,, \\ \mathrm{E}(\zeta_1 - \mathrm{E}\zeta_1)^{2\rho} &= 2a\,, \\ \mathrm{E}(\zeta_1 - \mathrm{E}\zeta_1)^{2\rho+1} &= 0\,, \\ \mathrm{E}(\zeta_2 - \mathrm{E}\zeta_2)^4 &= 4a(6a^2 + 6a + 1)\,, \\ \mathrm{E}\zeta_2^2 &= 4a + 1\,, \\ \mathrm{E}\zeta_2^3 &= 12a^2 + 12a + 1\,, \\ \mathrm{E}\zeta_2^4 &= 72a^2 + 24a^3 + 28a + 1\,, \end{aligned} \tag{29}$$

where $\rho \in \mathbb{N}_{\geq 1}$.

8 Code

This code is an original code written in R, free software environment for statistical computing and graphics. The code is convenient to use for prototyping solutions of the mathematical model described in this monograph or other models based on this model. Some parts of the code contain additional computational details too trivial for formal mathematical statements.

8.1 Algorithm C1. Computation of the conditional probability of the model with the restriction

Function *cp* computes the conditional probability given by (3) and lemmas on computational simplifications (9). The auxiliary function *expr* computes the corresponding parts given in (3) using computational lemmas (7) and (8).

```
expr = function(a,b,j,kappa,n) {
  if (n - kappa + j > 0) {
    rslt = choose(n,n - kappa + j)*a^(n - kappa + j)*
     choose(kappa - j, kappa - 2*j)*b^(kappa - 2*j)*(1 - a - b)^j
    if (is.nan(rslt) || is.infinite(rslt)) {
      rslt = a*(1 + (kappa - j)/(n - kappa + j))*expr(a,b,j,kappa,n-1)
    }
    rslt
  } else {
    if (n - kappa + j == 0) {
      rslt = choose(kappa - j, kappa - 2*j)*b^(kappa - 2*j)*(1 - a - b)^j
      if (is.nan(rslt) || is.infinite(rslt)) {
        if (kappa > 3*j) {
          rslt = prod(b*(1-a-b)*(1+(kappa-2*j)/(1:j)))*b^(kappa - 3*j)
        }
      }
      rslt
    } else {0}
  }
}

cp <- function(v,kappa,n) {
  # comment: kappa and n must be greater than zero;
  # n is the number of cells at time (t-1);
  # kappa is the number of cells at time t.
  a = v[1]; b = v[2]
  if (a == 0) {
    if (kappa >= n) {
      if (kappa <= 2*n) {
        rslt = choose(n,2*n - kappa)*b^(2*n - kappa)*(1 - b)^(kappa - n)
        if (is.nan(rslt) || is.infinite(rslt)) {
```

```
        if (3*n > 2*kappa) {
          rslt = prod(b*(1-b)*(1 + (2*n-kappa)/(1:(kappa - n))))*b^(3*n-2*kappa)
        } else {
          rslt = prod(b*(1-b)*(1 + (kappa - n)/(1:(2*n - kappa))))*(1-b)^(2*kappa - 3*n)
        }
      }
      rslt
    } else {0}
  } else {0}
} else {
  if (b == 0) {
    if (kappa <= 2*n) {
      if (1 - kappa %% 2) {
        rslt = choose(n,n - kappa/2)*a^(n - kappa/2)*(1 - a)^(kappa/2)
        if (is.nan(rslt) || is.infinite(rslt)) {
          if (n > kappa) {
            rslt = prod(a*(1-a)*(1 + (n-kappa/2)/(1:(kappa/2))))*a^(n-kappa)
          } else {
            rslt = prod(a*(1-a)*(1 + (kappa/2)/(1:(n-kappa/2))))*(1-a)^(kappa - n)
          }
        }
        rslt
      } else {0}
    } else {0}
  } else {
    if (a + b == 1) {
      if (kappa <= n) {
        rslt = choose(n,n - kappa)*a^(n - kappa)*b^kappa
        if (is.nan(rslt) || is.infinite(rslt)) {
          if (n > 2*kappa) {
            rslt = prod(a*b*(1 + (n - kappa)/(1:kappa)))*a^(n - 2*kappa)
          } else {
            rslt = prod(a*b*(1 + kappa/(1:(n-kappa))))*(1-a)^(2*kappa)
          }
        }
        rslt
      } else {0}
    } else {
      smm = 0
      if (kappa > n) {
        kappa = 2*n - kappa
        a = 1-a-b
      }
      for (j in 0:((kappa - kappa %% 2)/2)) {
        smm = smm + expr(a,b,j,kappa,n)
      }
      smm
    }
  }
}
}
}
```

8.2 Algorithm C2. Computation of the marginal probabilities of the model with the restriction

```
nT = 20 # comment: the maximal number of time steps.
vpsi = 64 # comment: the limit psi
sv = c(0.33,0.44) # simplex vector of probabilities requires to set only a and b.
```

```
D = mat.or.vec(nT + 1,vpsi + 1)
D[1,2] = 1 #P(kappa[0] = 1) = 1
for (t in 1:nT) {
    for (kappa in 0:min(vpsi,2^t)) {
        fp = 0
        for (iot in ((kappa + kappa %% 2)/2):min(vpsi,2^(t - 1))) {
            if (vpsi >= 2^t) {
              dlt = cp(sv,kappa,iot)
            } else {
              if (vpsi >= 2*iot) {
                dlt = cp(design[i,],kappa,iot)
              } else {
                prb = rep(0,vpsi + 1)
                for (ke in 0:vpsi) {
                  prb[ke + 1] = cp(design[i,],ke,iot) # comment: conditional probability
                }
                dlt = prb[kappa + 1]/sum(prb)
              }
            }
            fp = fp + dlt*D[t, iot + 1]
        }
        D[t + 1, kappa + 1] = fp
   }
}
}
```

Bibliography

[1] Mike West, 2013, Semester course “Probabilistic and statistical modeling” with corresponding lecture notes on time series and graphical models.

[2] Ksenia Kyzyurova, 2017, On Uncertainty Quantification for Systems of Computer Models, *Philosophy doctorate thesis*, Duke University.

[3] Ksenia Kyzyurova, 2018, Coupling computer models through linking their statistical emulators, *SIAM/ASA Journal on Uncertainty Quantification* **6**(3), 1151-1171.

[4] Ksenia Kyzyurova, 2019, Analysis of scientific computer models. Methodology in numerical simulator data analysis. `http://kseniak.ucoz.net`.

[5] Ksenia Kyzyurova, 2019, Calibration of mathematical computer models. `http://kseniak.ucoz.net`.