Posters
Discrete Bifurcation Analysis with Pithya

Nikola Beneš, Luboš Brim, Martin Demko, Matej Hajnal, Samuel Pastva, and David Šafránek

Systems Biology Laboratory, Faculty of Informatics, Masaryk University,
Botanická 68a, 602 00 Brno, Czech Republic
{xbenes3,brim,xdemko,xhajnal,xpastva,safranek}@fi.muni.cz

Bifurcation analysis is a central task of the analysis of parameterised high-dimensional dynamical systems that undergo transitions as parameters are changed. To characterise such transitions for models with many unknown parameters is a major challenge for complex, hence more realistic, models in systems biology. Its difficulty rises exponentially with the number of model components.

The classical numerical and analytical methods for bifurcation analysis are typically limited to a small number of independent system parameters. To address this limitation we have developed a novel approach to bifurcation analysis, called discrete bifurcation analysis, that is based on a suitable discrete abstraction of the given system and employs model checking for discovering critical parameter values, referred to as bifurcation points, for which various kinds of behaviour (equilibrium, cycling) appear or disappear. To describe such behaviour patterns, called phase portraits, we use a hybrid version of a CTL logic augmented with direction formulae.

Technically, our approach is grounded in a novel method of parameter synthesis from temporal logic formulae using symbolic model checking and implemented in a new high-performance tool Pithya\(^1\) [1]. Pithya itself implements state-of-the-art parameter synthesis methods. For a given ODE model, it allows to visually explore model behaviour with respect to different parameter values. Moreover, Pithya automatically synthesises parameter values satisfying a given property. Such property can specify various behaviour constraints, e.g., maximal reachable concentration, time ordering of events, characteristics of steady states, the presence of limit cycles, etc. The results can be visualised and explored in a graphical user interface.

We demonstrate the method on a case study taken from biology describing the interaction of the tumour suppressor protein pRB and the central transcription factor E2F1 [3]. This system represents an important mechanism of a biological switch governing the transition from \(G_1\) to \(S\) phase in the mammalian cell cycle. In the \(G_1\)-phase the cell makes an important decision. In high concentration levels, \(E2F1\) activates the phase transition. In low concentration of \(E2F1\), the transition to \(S\)-phase is rejected and the cell avoids division.

---

\(^1\) http://biodivine.fi.muni.cz/pithya/.

This work has been supported by the Czech National Infrastructure grant LM2015055 and by the Czech Science Foundation grant GA15-11089S.

© Springer International Publishing AG 2017
DOI: 10.1007/978-3-319-67471-1
References


We propose a simple parallel stochastic dynamics for understanding random association cluster formations of $2n = 40$ *Mus musculus domesticus* bivalents during pachytene in early prophase and provide statistically optimized parameters for ensuring adequate fitting of the model with available experimental data [2]. This work represents a continuation of the discrete dynamical approach started in [2, 3] while modeling randomness for chromosome associations in $2n = 40$ - *Mus m. domesticus* spermatocytes. We focus on pachytene in prophase I (see [1]).

During pachytene, at the prophase stage of meiosis, the homologous chromosomes synapse along a proteinacious structure, called synaptonemal complex (SC), thus enabling recombination between them, a process that produces genetic variation. The synapsed chromosomes are called bivalents and are found to be attached to the nuclear envelope by both their ends, being able to move or glide upon the internal surface of it. Chromosomal bivalent’s associations are said to be given by intersecting domains of *constitutive pericentromeric heterochromatin* (CPCH’s), which are known to create rich dynamic and diverse scenarios via the participating elements. These are triggered by the corresponding intersection domains of CPCH located at the short arms of each bivalent, but also by the associated convergence of the rest of the constituent chromatin along them. These structures are revealed by means of squashes (or spreads), in which the nuclear envelope is removed and the spermatocyte’s nucleus content is projected to a flat surface.

Data from 400 pachytene spermatocyte spreads of $2n = 40$ *Mus domesticus* treated by immunocytochemical techniques taken from [2] is used for contrasting theoretical results: we model the spermatocyte’s nucleus as an (almost) six regular graph, which ensure maximal connectivity for the nodes. They represent the positions of the bivalents attached to the nuclear envelope.
Upon this discrete surface, the SC evolution follows a parallel rewriting rule as in Fig. 1. Here, the bivalent’s structures are represented as attributes of the vertices: The synaptonemal complex SC attached to the envelope and a random neighborhood of vertices for the CPCH. Pathwise connected domains of overlapping CPCH are considered to build an association cluster between the corresponding bivalents and we describe and analyze their statistical and dynamical distribution.

The model can now be used for interrogating different phenomena associated to the superposition of chromatin domains of the bivalents during pachytene, as well as providing a theoretical description of the kind of randomness involved in these phenomena (Fig. 2).

Undoubtedly, a model-theoretical approach to the general principles behind bivalent’s associations in prophase meiotic nuclei, as well as precising the type of randomness being at play at this stage could bring us also a step closer to a better understanding of the different chromosome combinations present in the gametes. Since these associations and combinations persist until the meiotic divisions, the chromosomal associations as described here necessarily leave some imprint in the chromosomal sets passed on to gametes and hence their importance to evolution.

Acknowledgements. This work has been partially supported by Chilean MINEDUC Grant MECE-SUP 2016-2017 and LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d’avenir.
References

Non-disjoint Clustered Representation for Distributions over a Population of Cells

Matthieu Pichéne\textsuperscript{1}, Sucheendra Palaniappan\textsuperscript{2}, Eric Fabre\textsuperscript{1}, and Blaise Genest\textsuperscript{3}

\textsuperscript{1} Inria, Team SUMO, Rennes, France
matthieu.pichene@inria.fr
\textsuperscript{2} The Systems Biology Institute, Tokyo, Japan
\textsuperscript{3} CNRS, IRISA, Rennes, France

1 Motivation

We consider a large homogenous population of cells, where each cell is governed by the same complex biological pathway. A good modeling of the inherent variability of biological species is of crucial importance to the understanding of how the population evolves. In this work, we handle this variability by considering multivariate distributions, where each species is a random variable. Usually, the number of species in a pathway -and thus the number of variables- is high. This appealing approach thus quickly faces the curse of dimensionality: representing exactly the distribution of a large number of variables is intractable.

To make this approach tractable, we explore different techniques to approximate the original joint distribution by meaningful and tractable ones. The idea is to consider families of joint probability distributions on large sets of random variables that admit a compact representation, and then select within this family the one that best approximates the desired intractable one. Natural measures of approximation accuracy can be derived from information theory. We compare several representations over distributions of populations of cells obtained from several fine-grained models of pathways (e.g. ODEs). We also explore the interest of such approximate distributions for approximate inference algorithms \cite{1, 2} for coarse-grained abstractions of biological pathways \cite{3}.

2 Results

Our approximation scheme is to drop most correlations between variables. Indeed, when many variables are conditionally independent, the multivariate distribution can be compactly represented. The key is to keep the most relevant correlations, evaluated using the mutual information (MI) between two variables.

The simplest approximation is called fully factored (FF), and assumes that all the variables are independent. It leads to very compact representation and fast computations, but it also leads to fairly inaccurate results as correlations between variables are entirely lost, even for highly correlated species (MI = 0.6).
Alternately, one can preserve a few of the strongest correlations, selected using MI, giving rise to a set of disjoint clusters of variables. For efficiency reason, we used clusters of size two. This model was able to capture some of the most significant correlations between pairs of variables (representing around 30\% of the total MI), but dropped significant ones (MI = 0.2).

A better trade-off between accuracy and tractability was obtained by using non-disjoint clusters of two variables, structured as a tree, called the tree-clustered approximation (TCA). The approximated joint distribution is fully determined by the marginals over each selected cluster of 2 variables. This gives a compact representation (<800 values in our experiments). Further, any marginal over \( k \) out of \( n \) total variables can be computed with time complexity \( O(nv^{k+1}) \), where each variable can take \( v \) possible values. Last, a tractable algorithm [4] allows to compute the best approximation of any distribution by a tree of clusters. TCA succeeded in capturing most correlations between pairs of variables (representing around 70\% of the total MI), losing no significant ones (MI < 0.1).

Regarding inference, FF, disjoint clusters and TCA were compared to Hybrid FF (HFF) [2]. In short, HFF preserves a small number of joint probabilities of high value (called spikes), plus an FF representation of the remaining of the distribution. The more spikes, the more accurate the approximation, and the slower HFF inference. Overall, TCA is very accurate, while HFF generates sizable errors, even with numerous spikes (32k). Further, TCA is faster than HFF, even with few spikes (3k). FF and disjoint-clusters are even faster (1 to 2 order of magnitudes) than TCA, but the accuracy of both remains problematic.

3 Perspectives

We now aim at modeling and studying a tissue, made of tens of thousands of cells. In this context, capturing the inherent variability of the population of cells is crucial. In order to study multi-scale systems in a tractable way, we advocate a two-step approach: Firstly, abstract the low level model of the pathway of a single cell into a stochastic discrete abstraction, e.g. using [3]. Secondly, use a model of the tissue, which does not explicitly represent every cell but qualitatively explains how the population evolves. In this way, one need not explicitly represent the concentration of each of the tens of thousands of cells, but rather only keep one probability distribution.

Acknowledgement. This work was partially supported by ANR-13-BS02-0011-01 STOCH-MC.
References

Effects of the Dynamics of the Steps in Transcription Initiation on the Asymmetry of the Distribution of Time Intervals Between Consecutive RNA Productions

Sofia Startceva\textsuperscript{1}, Vinodh Kumar Kandavalli\textsuperscript{1}, Ari Visa\textsuperscript{2}, and Andre S. Ribeiro\textsuperscript{1(\textsuperscript{ESD})}

\textsuperscript{1} Laboratory of Biosystem Dynamics, BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, 33101 Tampere, Finland
andre.ribeiro@tut.fi

\textsuperscript{2} Signal Processing Unit, Faculty of Computing and Electrical Engineering, Tampere University of Technology, 33101 Tampere, Finland

Abstract. Asymmetries in the distribution of time intervals between consecutive RNA productions from a gene can play a critical role in, e.g., allowing/preventing the RNA and, thus, protein numbers to cross thresholds involved in gene network decision making. Here, we use a stochastic, multi-step model of transcription initiation, with all rate constants empirically validated, and explore how the kinetics of its steps affect the temporal asymmetries in RNA production, as measured by the skewness of the distribution of intervals between consecutive RNA productions in individual cells. From the model, first, we show that this skewness differs widely with the mean fraction of time that the RNA polymerase spends in the steps preceding open complex formation, while being independent of the mean transcription rate. Next, we provide empirical validation of these results, using qPCR and live, time-lapse, single-molecule RNA microscopy measurements of the transcription kinetics of multiple promoters. We conclude that the skewness in RNA production kinetics is subject to regulation by the kinetics of the steps in transcription initiation and, thus, evolvable.

Keywords: Transcription initiation · Asymmetries in RNA production · Stochastic models · Single-RNA measurements

Gene expression regulation in bacteria occurs mostly in transcription initiation [1]. In \textit{Escherichia coli}, this process is sequential [2], starting with an RNA polymerase (\(R\)) binding to an active promoter (\(P_{ON}\)) and forming a closed complex (\(RP_{cc}\)). Next, the open complex (\(RP_{oc}\)) forms. Relevantly, the subsequent steps of RNA elongation [3], termination, and RNA and \(R\) release are much faster. Thus, dynamically, transcription can be approximately modeled as:

\[
R + P_{ON} \xrightarrow{k_e} RP_{cc} \xrightarrow{k_e} RP_{oc} \xrightarrow{\infty} P_{ON} + R + \text{RNA}
\]

© Springer International Publishing AG 2017
DOI: 10.1007/978-3-319-67471-1
Here, RNA production kinetics is controlled by $k_{cc}$ and $k_{oc}$. The probability density function (pdf) of the distribution of intervals between transcription events is the convolution of their pdfs:

$$f_D(t) = \frac{k_{cc}}{k_{cc} - k_{oc}} \left(e^{-k_{oc}t} - e^{-k_{cc}t}\right).$$

To measure asymmetries in this distribution, we use skewness, $S = \frac{m_3}{m_3^2}$, where $m_r = \frac{1}{n} \sum (x_i - \bar{x})^r$ [4]. We estimate the sample skewness $S_s = \sqrt{\frac{n(n-1)}{n-2}} \cdot S$, where $n$ is the sample number [5]. To obtain confidence boundaries for $S_s$ we use non-parametric bootstraps as in [6].

In (1), $k_{cc}$ is the inverse of the mean time for $R$ to bind the promoter and complete a closed complex ($\tau_{cc}$), while $k_{oc}$ is the inverse of the mean time for an open complex to form ($\tau_{oc}$). The mean time between transcription events: $\Delta t = \tau_{cc} + \tau_{oc}$.

To validate the model predictions of skewness, we collected empirical data for $\Delta t$ and $\tau_{cc}/\Delta t$ for various promoters ($P_{TetA}$, $P_{BAD}$, $P_{Lac-ara-1}$, and $P_{Lac-ara-1}$ under oxidative stress) [7–9] (Fig. 1). Next, given the mean $\Delta t$ of each promoter, we varied $\tau_{cc}/\Delta t$ (from 0 to 1) while maintaining $\Delta t$ constant. Then, for each value of $\tau_{cc}/\Delta t$, we calculated $S$ from the pdf of the distribution of intervals between transcription events (solid line, Fig. 1). Interestingly, we observed that $S$ is independent of the mean value of $\Delta t$. Finally, from Fig. 1, we find that the model predictions of $S$ fit the empirical data.

![Fig. 1. Predicted skewness of $\Delta t$ distributions with given $\tau_{cc}/\Delta t$ (solid line) and sample skewness of the empirical $\Delta t$ distributions (with 95% confidence intervals) for the studied promoters. For each promoter, 100 or more $\Delta t$ intervals were extracted from a total of 100 or more cells.](image)

Importantly, as $S$ is tunable by $\tau_{cc}$ and $\tau_{oc}$, which are sequence dependent and subject to regulation, we expect it to be evolvable and adaptive to environment shifts.

References

Author Index

Abbas, Houssam 23
Barnat, Jiří 40
Bartocci, Ezio 23
Beneš, Nikola 40, 319
Berrios, Soledad 321
Bernot, Gilles 196
Biane, Célia 57
Boba, Patrick 285
Bournez, Olivier 108
Brim, Luboš 40, 319

Camporesi, Ferdinanda 291
Carcano, Arthur 74
Comet, Jean-Paul 196
Coquet, Jean 91
Cornec, Yves-Stan Le 3
Cummings, Bree 300

Dameron, Olivier 91
Delaplace, Franck 57
Demko, Martin 40, 319

Echahed, Rachid 321
Fabre, Eric 324
Fages, François 74, 108
Faverney, Christine Risso-de 196
Fenner, Julio López 321
Feret, Jérôme 291

Gedeon, Tomas 300
Genest, Blaise 324
Giehr, Pascal 160
Grousi, Radu 23

Haar, Stefan 179
Hajnal, Matej 40, 319
Hamacher, Kay 285
Harker, Shaun 300
Harmer, Russ 3
Hasenauer, Jan 128

Kandavalli, Vinodh Kumar 327
Klinger, Emmanuel 128
Knapp, Merrill 249

Le Guludec, Guillaume 108
Legagneux, Vincent 91
Légaré, Sébastien 3
Liang, Kai-Wen 145
Liu, Kin Sum 214
Lück, Alexander 160
Lý, Kim Quyền 291

Maignan, Aude 321
Mandon, Hugues 179
Miraglio, Benjamin 196
Misaïkov, Konstantin 300
Miskov-Zivanov, Natasa 145

Oshurko, Ievgeniia 3

Paolletti, Nicola 214
Pastva, Samuel 40, 319
Paulévé, Loïc 179, 309
Palaniappan, Sucheendra 324
Pichené, Matthieu 324
Pouly, Amaury 108

Ramanathan, R. 265
Ravichandran, Divyaa 145
Ribeiro, Andre S. 327
Rodionova, Alena 23

Šafránek, David 40, 319
Schwieger, Robert 233
Siebert, Heike 233
Smolka, Scott A. 23, 214
Soliman, Sylvain 74
Spirtes, Peter 145
Startceva, Sofia 327
Talcott, Carolyn 249
Telmer, Cheryl 145
Theret, Nathalie 91
Thiagarajan, P.S. 265
Visa, Ari 327

Walter, Jörg 160
Wang, Qinsi 145
Wolf, Verena 160
Wong, Weng-Fai 265
Zhou, Jun 265