Identification of individual-based models of growth and gene expression dynamics from population-snapshot data

Funded PostDoc proposal (15 months), Université Grenoble Alpes.
Desired start date: March 1st, 2023.
Contacts: Aline Marguet (aline.marguet@inria.fr),
Adeline Leclerq-Samson (adeline.leclercq-samson@univ-grenoble-alpes.fr)

Keywords: Branching processes, statistical inference, sensitivity analysis, least-squares estimation, ABC methods, optimal transport, single-cell measurements, microbial dynamics.

Introduction. Modern experimental technologies allow one to monitor gene expression in microbial cells at the single-cell level. This has allowed to quantitatively assess the variability of gene expression across cells of a genetically identical population. Developing models and inference methods that can explain the variability observed in single-cell datasets is a fundamental step toward the understanding of origins and consequences of heterogeneity of microbial populations [3, 5].

Several methods have been proposed for the inference of stochastic models of gene expression from population-snapshot data, that is, quantitative single-cell measurements of gene expression and other properties (cell size) in cell samples drawn from a growing population at different times [6, 2]. Common assumptions are that individual cell dynamics are statistically independent outcomes of a same stochastic process, and that dilution effects associated with individual cell growth (in the sense of increase of cell volume) can be quantified in terms of the growth rate of the whole population.

However, growth rate is also heterogeneous across cells, which introduces an additional source of variability in individual gene expression. In presence of inheritance from mother to daughter cells, faster growing cells may divide faster and be increasingly represented in the population. Therefore, in order to investigate the effects of growth rate variability on the statistics of gene expression, the assumption of statistically independent cells poses problems.

Objectives. The aim of this project is to develop methods for the inference of individual-based gene expression models from population-snapshot data. Often adopted to describe growing populations of individuals,the individual-based modelling framework provides an ideal tool to describe individual cell growth and gene expression dynamics as well as inheritance of individual traits [4, 1]. Yet inference tools in presence of growth variability are lacking. Based on a given individual-based model comprising (A) a parametric model of individual cell growth rate and its stochastic inheritance at cell division, and (B) a parametric model of (stochastic) gene expression dynamics, the proposal is arranged into the following tasks:

1. Implement and simulate the dynamics of the population distribution of gene expression (and cell size), by means of stochastic simulation of branching processes or numerical solution of corresponding (partial) differential equations, for different hypothetical growth conditions / gene expression stimuli;

2. Perform sensitivity analysis of population distributions to variations of parameters / identifiability of the growth rate model (A) and of the gene expression model (B), based on local analysis and/or simulation with the tools developed in 1;

3. Develop methods for the inference of model parameters from population-snapshot data. The methods foreseen are based on matching the distributions of gene expression / cell size at different times that are provided by the data with corresponding model-predicted distributions (least-squares, ABC, optimal transport, etc.). These parameter-dependent model predictions shall be obtained with the methods developed in point 1. Specifically, depending on the identifiability results in 2, we aim to develop, implement and test methods to identify (i) parameters of (A) knowing (B), (ii) parameters of (B) knowing (A), (iii) parameters of both (A) and (B);

4. Apply the methods on simulated datasets and/or on real experimental datasets.

Skills. The candidate is expected to hold a Ph.D. in statistics, applied mathematics or other relevant field, as well as familiarity with or strong interest for biological systems. He/she is expected to be proactive and open to collaboration in an interdisciplinary and international working context.

Context. The project will be developed in collaboration with A. Marguet and E. Cinquemani (Systems Biology team MICROCOSME, Inria de l'Université Grenoble Alpes), A. Samson-Leclerq (Laboratoire Jean Kuntzmann, Université Grenoble Alpes) and L. Coquille (Institut Fourier, Université Grenoble Alpes). It will be co-localized at Inria and LJK.

References

[1] A. Marguet, M. Lavielle, and E. Cinquemani. Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data. *Bioinformatics*, 35(14):i586–i595, 2019.

[2] B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(1):318, Jan. 2009.

[3] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26, 2008.

[4] P. Thomas and V. Shahrezaei. Coordination of gene expression noise with cell size: analytical results for agent-based models of growing cell populations. *Journal of The Royal Society Interface*, 18(178):20210274, 2021.

[5] S. Waldherr. Estimation methods for heterogeneous cell population models in systems biology. *Journal of The Royal Society Interface*, 15(147):20180530, 2018.

[6] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences*, 109(21):8340–8345, May 2012.