Methods xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Methods

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

Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population

L. Bronstein [a], C. Zechner [b], H. Koeppl [a,*]

[a] Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Darmstadt, Germany
[b] Department of Biosystems Sciences and Engineering, ETH Zürich, Basel, Switzerland

ARTICLE INFO

Article history:
Received 9 February 2015
Accepted 10 May 2015
Available online xxxx

Keywords:
Bayesian inference
Stochastic models
Cell-to-cell variability

ABSTRACT

Single-cell experimental techniques provide informative data to help uncover dynamical processes inside a cell. Making full use of such data requires dedicated computational methods to estimate biophysical process parameters and states in a model-based manner. In particular, the treatment of heterogeneity or cell-to-cell variability deserves special attention. The present article provides an introduction to one particular class of algorithms which employ marginalization in order to take heterogeneity into account. An overview of alternative approaches is provided for comparison. We treat two frequently encountered scenarios in single-cell experiments, namely, single-cell trajectory data and single-cell distribution data.
© 2015 Published by Elsevier Inc.

1. Introduction

Single-cell techniques provide a more direct access to the inner workings of a cell than classical bulk measurements. Although bulk measurements are sometimes considered advantageous due to the inherent averaging and filtering of unwanted fluctuations, it is becoming increasingly clear that the actual mechanisms are occluded by such averaging (see Fig. 1 for an illustrative example).

The mathematical basis for this observation is that the mean behavior of the bulk can generally not be traced back to a single mean mechanism (e.g., a mean threshold of a switch). Moreover, even if such a mean mechanism exists, it would not give rise to the mean behavior observed at the bulk due to the presence of non-linearities. This discrepancy stems from the non-commutativity between the averaging operation and the finite-time evolution operator, due to the nonlinearity of the latter for almost all cases. In particular, this even persists for perfectly linear reaction kinetics.

Experimental techniques for single-cell analysis advance rapidly and provide data of increasing accuracy and dimensionality [1]. For instance, while classical fluorescent-based flow cytometry or FACS (fluorescence-activated cell sorting) is limited to a handful of channels [2], mass cytometry [3] with rare-earth labeled anti-bodies go to 50 and more simultaneous channels while attaining comparable throughput in terms of number of cells. Although at significantly lower throughput single-cell transcriptome analysis by RNA-Seq [4] provides genome-wide estimates of RNA abundances on the single-cell level. Specific RNAs can be counted in fixed cells through fluorescence *in situ* hybridization (FISH) [5,6], while RNA-labeling through bacterial coat proteins yields information about transcription and RNA abundance of a specific gene in living cells [7–9]. Techniques such as characterizing the expression of a gene using fluorescent reporter proteins, fusion proteins or luciferases share with all live-cell techniques that they only provide low dimensional readouts. That is, similar to flow cytometry they are restricted to a few channels. Live-cell techniques however provide a qualitatively distinct feature in so much that a single cell can be monitored over the course of time. Such single-cell trajectory data contains temporal correlation on the single-cell level that is otherwise not accessible. For instance, techniques involving fixed cells, or flow cytometry based techniques lack this correlative information. Note that this also applies to live-cell flow cytometry protocols where the same cells are measured at successive time points because the cell identity is lost.

In this article, we give an overview of inference algorithms that can extract the information contained in the results of the experimental techniques described above. The term *inference* here refers to the process of using information available to us in order to quantify our state of knowledge about formulated hypotheses or unknown parameters. Doing this in a systematic and principled way involves various mathematical ideas. These have to be introduced before we come to actual inference algorithms. We begin by describing the mathematical models used for biomolecular reaction networks, the biologically important notion of cell-to-cell variability, general aspects of Bayesian inference and the mathematical modeling of measurement procedures in Section 2. The concept of

* Corresponding author.
E-mail address: heinz.koeppl@bcs.tu-darmstadt.de (H. Koeppl).

http://dx.doi.org/10.1016/j.ymeth.2015.05.012
1046-2023/© 2015 Published by Elsevier Inc.

ARTICLE IN PRESS

Key notation

$\mathcal{S}$	intrinsic parameters.	y_n^m	single-cell distribution measurement of the m-th cell at the n-th time point.
$\mathcal{Z}$	extrinsic factors.	x_n^m	underlying true state for single-cell distribution data.
α	extrinsic statistics.	$\mathbf{y}_n^m$	single-cell trajectory measurement of the m-th cell at the n-th time point.
ν_r	stoichiometric change vector of reaction r.	$\mathbf{x}_n^m$	underlying true state for single-cell trajectory data.
$h_r(c_r, x)$	propensity of reaction r in state x.	$\mathcal{D}$	complete set of measured data.
$X(t)$	process state at time t.		
$\mathbf{X}_{[0,t]}$	complete path of the process over the interval $[0, t]$.		

marginalization, key to this article, is explained in Section 3. Actual inference algorithms are described in Section 4.

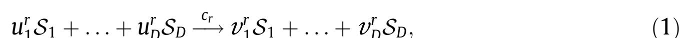
We should state clearly that the goal of this article is not to give a complete review of all available techniques. Instead, we focus on methods which are able to deal with cell-to-cell variability, and do so by employing the concept of marginalization. In this sense, the selection of topics of this review is obviously biased towards the work of some of the authors.

2. Models and problem statement

In this section preliminary concepts needed for the description of inference algorithms are discussed. The mathematical models that are commonly used for reaction networks are explained in Section 2.1, and the modeling of heterogeneity in Section 2.2. This is followed by a description of the experimental procedures that we consider in Section 2.3. We explain the general principles behind Bayesian inference, and how they relate to biomolecular reaction kinetics, in Sections 2.4 and 2.5, and introduce the issue of non-identifiability in Section 2.6. We begin with the mathematical models for biomolecular reaction processes.

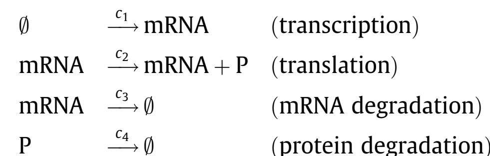
2.1. Markovian population models

Throughout this review we consider a process abstraction that operates on the intracellular population levels of involved molecular species and on their interaction in terms of biomolecular transformations. Hence, we assume the absence of any truly spatial effects for which spatial models such as Brownian dynamics would be necessary [10,11]. Time invariant spatial arrangements of compartments for each of which the well-mixed assumption holds, can however be modeled by such population models when augmented by suitable transport reactions. Subsequently, we assume that the process of interest involves D molecular species $\mathcal{S}_1, \ldots, \mathcal{S}_D$ that interact through R different reactions. The r-th reaction is described by its stoichiometry



$$u_1^r \mathcal{S}_1 + \ldots + u_D^r \mathcal{S}_D \xrightarrow{c_r} \nu_1^r \mathcal{S}_1 + \ldots + \nu_D^r \mathcal{S}_D, \quad (1)$$

where u_d^r and ν_d^r denote the integer-valued stoichiometric substrate and product coefficients, respectively, of the d-th species and c_r is the real-valued reaction rate constant. The stoichiometric change vector $\nu_r \equiv (\nu_1^r - u_1^r, \ldots, \nu_D^r - u_D^r)$ describes the net change in the process state when the reaction occurs. Throughout this article, we use a simple model of gene expression, given by



$$\begin{array}{lll} \emptyset & \xrightarrow{c_1} \text{mRNA} & \text{(transcription)} \\ \text{mRNA} & \xrightarrow{c_2} \text{mRNA} + \text{P} & \text{(translation)} \\ \text{mRNA} & \xrightarrow{c_3} \emptyset & \text{(mRNA degradation)} \\ \text{P} & \xrightarrow{c_4} \emptyset & \text{(protein degradation)} \end{array}$$

as a running example. Here we have two species and four reactions with stoichiometric change vectors $\nu_1 = (1, 0)$, $\nu_2 = (0, 1)$, $\nu_3 = (-1, 0)$ and $\nu_4 = (0, -1)$, where we take the mRNA to be the first species and the protein to be the second.

The inference algorithm and consequently, the result it returns, depend on the way we translate those reactions into a mathematical model. Several such models are in common use in computational biology and can somewhat be organized into a hierarchy (see Fig. 2), facilitating comparison and understanding. We begin by describing the model at the top of the hierarchy, called a continuous-time Markov chain (CTMC), instantiating a general class of models that are defined on the discrete state-space of copy numbers for all involved species. Accordingly, the state at time t of the above reaction system, denoted by $X(t)$, is a vector of length D with integer-valued entries that computes to

$$X(t) = X(0) + \sum_{r=1}^{R} N_r(t) \nu_r, \quad (2)$$

where $N_r(t)$ denotes the number of reactions of type r up to time t. In our running example, we have $X(t) = (X_1(t), X_2(t))$ where X_1 is the abundance of mRNA and X_2 the abundance of protein. For a continuous-time Markov chain it moreover holds that at any given time t with $X(t) = x$, the probability of reaction r to occur in the time interval between t and $t + \Delta t$ is approximately $h_r(c_r, x)\Delta t$, for sufficiently small Δt. Here $h_r(c_r, x)$ is the propensity or hazard of the reaction, which is often assumed to be of the form $c_r g_r(x)$, where g_r is some function of the current state x. In particular, the probability of a reaction at any given time is independent of the history of the process, as it only depends on the current state. This is the

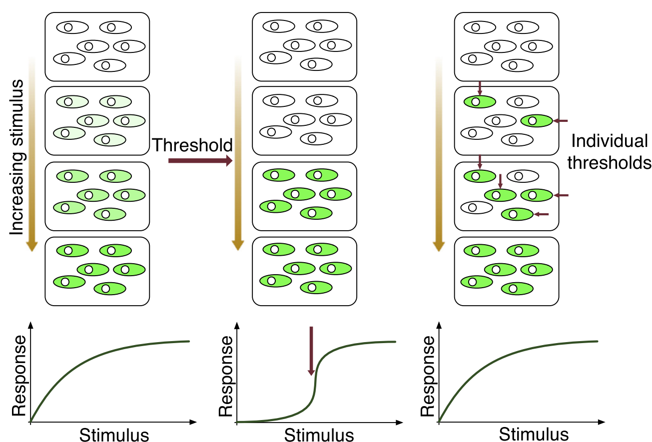


Fig. 1. Determining the dose response of a cell population. Cells respond to the stimuli in a graded, homogeneous manner (left), in a homogeneous switch-like manner (middle) or in a heterogeneous switch-like manner (right); while the actual mechanism on the single-cell level is very different between left and right, the difference cannot be detected in the dose–response curve of the bulk.

Please cite this article in press as: L. Bronstein et al., Methods (2015), http://dx.doi.org/10.1016/j.ymeth.2015.05.012

Download English Version:

https://daneshyari.com/en/article/8340564

Download Persian Version:

https://daneshyari.com/article/8340564

Daneshyari.com