Statistical Model Checking Based Calibration and Analysis of Bio-pathway Models*

Sucheendra K. Palaniappan[1], Benjamin M. Gyori[2], Bing Liu[3], David Hsu[1,2], and P.S. Thiagarajan[1,2]

[1] School of Computing, National University of Singapore, 117417, Singapore
[2] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 117417, Singapore
[3] Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract. We present a statistical model checking (SMC) based framework for studying ordinary differential equation (ODE) models of bio-pathways. We address cell-to-cell variability explicitly by using probability distributions to model initial concentrations and kinetic rate values. This implicitly defines a distribution over a set of ODE trajectories, the properties of which are to be characterized. The core component of our framework is an SMC procedure for verifying the dynamical properties of an ODE system accompanied by such prior distributions. To cope with the imprecise nature of biological data, we use a formal specification logic that allows us to encode both qualitative properties and experimental data. Using SMC, we verify such specifications in a tractable way, independent of the system size. This further enables us to develop SMC based parameter estimation and sensitivity analysis procedures. We have evaluated our method on two large pathway models, namely, the segmentation clock network and the thrombin-dependent MLC phosphorylation pathway. The results show that our method scales well and yields good parameter estimates that are robust. Our sensitivity analysis framework leads to interesting insights about the underlying dynamics of these systems.

1 Introduction

Biochemical networks–often called bio-pathways–govern a variety of cellular functions. Their malfunctioning can lead to major diseases [1]. Thus it is important to understand their dynamics using mathematical models [2]. However, building and analyzing such models poses considerable challenges. In this paper, we address the particular challenge of accounting for variable behavior across individual cells. A natural way to cater for this is to use a probabilistic system model such as continuous time Markov chains (CTMCs) [3]. However, such models typically track the occurrences of individual reactions. Hence for pathways of realistic size, calibrating these models using experimental data and analyzing them using stochastic simulations is very difficult. The alternative is to use ordinary differential equations (ODEs) to capture the dynamics. This approach is often computationally more tractable, although it requires that the number of molecules of each type involved in the pathway be abundantly present [4]. In this paper our focus

* This research was partially supported by the Singapore MOE ARC grant MOE2011-T2- 2-012.

A. Gupta and T.A. Henzinger (Eds.): CMSB 2013, LNBI 8130, pp. 120–134, 2013.

is on accounting for cell-to-cell variability in the setting of ODE based models. Specifically, our main contribution is a statistical model checking (SMC) based framework, using which a system with such variability can be efficiently calibrated and analyzed.

Variability in a population of cells has at least two major causes. First, as shown in [5], differences in the initial concentrations of proteins are the primary source of variability in response to external stimuli. Second, due to differing internal and external conditions among cells, the values of kinetic rate constants also vary across cells [6, 7]. In our ODE setting the variables will represent the concentrations of the biochemical species (typically proteins) in the pathway, and hence the initial concentrations of these species will constitute the initial values of the variables. Further, the parameters appearing in the equations will consist of the kinetic rate constants governing the reactions. Thus we can capture cell-to-cell variability in the behavior of the bio-pathway by studying the ODE dynamics across a range of values for the initial concentrations and kinetic rate constant values. We do this in a probabilistic setting by assuming initial probability distributions (usually uniform) over an interval of values for the initial concentrations and rate constants. We then show that the resulting space of trajectories can be used to construct a natural probability measure space if the vector field defined by the ODE system is continuously differentiable. In our setting this requirement is easily met.

To analyze the ODE system, we first formalize properties using our specification logic and decide a corresponding confidence level (probability) with which we wish to assess them. Consequently, an SMC procedure –which poses the problem as a hypothesis test– is used to decide approximately, but with statistical guarantees, whether the properties are satisfied with the desired probability. SMC continues to sample and verify trajectories from the ODE system until a decision can be made. It is well-established that SMC is efficient since its complexity does not depend on the size of the system. Moreover, posing the problem as a sequential hypothesis test reduces the overall number of samples needed to make a decision [8]. These components form a principled method for analyzing the dynamics of a bio-pathway in the presence of dynamic variability across a population of cells.

To demonstrate the applicability of our approach, we develop an SMC based parameter estimation method. The unknown model parameters usually consist of initial concentrations and kinetic rate constants. Here, for convenience, we shall assume all the initial concentrations are known but that their nominal values can vary over a cell population. The parameter estimation procedure searches through the value space of the unknown parameters to determine the "best" combination of values that can explain the given data and predict new behaviors [9]. The key step in this procedure is to determine the fitness-to-data of the current set of parameter values. We use our specification logic to encode both experimental time series data and known qualitative trends concerning the dynamics of the pathway. We then use our SMC procedure to determine the goodness of the given set of parameter values, while taking into account that these values can fluctuate across the population of cells that the data is based on. Subsequently, we use a global optimization strategy known as SRES [10] to choose a new set of candidate parameter values according to the SMC based score assigned to the current set.

An important analysis task to be performed on the model is quantifying the influence of different parameters on the model dynamics. The information gained from such