Generalized Method of Moments improves parameter estimation in biochemical signaling models of time-stamped single-cell snapshot data

Wu John[1], Stewart William CL[2], Jayaprakash Ciriyam[3], and Das Jayajit[4].

***Abstract*— Ordinary differential equations (ODEs) are commonly used to model the sub-cellular dynamics of average values of proteins and mRNAs. However, parameter estimation in ODEs can still be challenging since the magnitudes of single-cell abundances can vary widely between proteins/mRNA. We employ Generalized Method of Moments (GMM) and Particle Swarm Optimization (PSO) to address the above challenges in mechanistic modeling of signaling kinetics data using synthetic and CYTOF data. Our results demonstrate substantial improvements for modeling data from single-cell cytometry and RNA-seq experiments.**

I. Motivation

ODEs are used to describe the deterministic dynamics of average concentrations, and estimation of model parameters from experimental data is an important step towards building biological models. The task of estimating parameters is challenging due to a variety of reasons, one of which is that the number of model parameters is larger than that of the available data. Recent developments of experimental techniques for the longitudinal measurements of transcripts and proteins at the single cell level e.g. single-cell RNA seq, CyTOF for proteins appear to alleviate this problem. For any time t, the cell-to-cell variability in the copy numbers of proteins and RNA arises from two sources: variation present at the pre-stimulus state (*aka* extrinsic noise), and variation that arises from the stochasticity of biochemical reactions (*aka* intrinsic noise). Especially, in developing models to describe signaling kinetics when the protein abundances are large, extrinsic noise is known to play a significant role. Thus, the single-cell cytometry data provide information about the probability distribution of different proteins due to extrinsic noise across cells. This leads to the availability of more observables, e.g., the various moments of the protein numbers, than the number of parameters in an ODE model. However, single cell abundances of different protein species can vary by orders of magnitudes, thus mean values and higher moments of these abundances can differ by many orders of magnitudes. Hence, the challenge here is to estimate parameters reliably given these large variations.

II. Results

We employ Generalized Method of Moments (GMM), a widely used method in econometrics, and Particle Swarm Optimization to address the above challenges. Using synthetic data from linear and non-linear models, we show that the proposed method improves parameter estimation and enables construction of approximate confidence intervals. We apply our parallelized software CyGMM (https://github.com/jhnwu3/CyGMM) to estimate parameters in linear and nonlinear ODE models for synthetic data, and published longitudinal single-cell cytometry data for CD8+ T cells. Our results demonstrate CyGMM provides substantial improvements for parameter estimation in ODE models of signaling kinetics using single cell cytometry data.

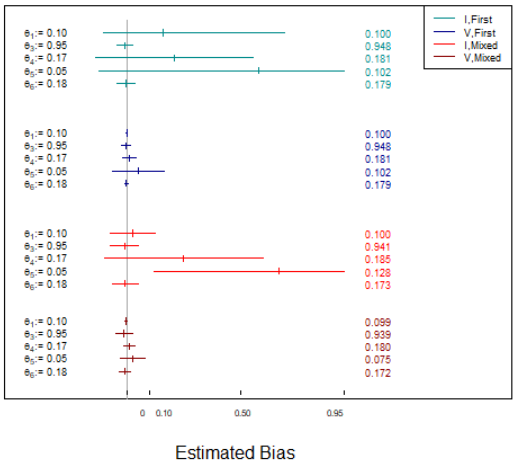

Fig. 1: Near optimal weights improve estimation, and higher moments can reduce bias. Four GMM estimators of the nonlinear model are shown. Two estimators consider differences in means (light blue and dark blue), and two other estimators consider differences in first, second, and mixed moments (light red and dark red). From the legend, I denotes equal weights and V the unequal (optimal) weights. Simulation truth and the average estimate across replicates are shown in the left and right margins.

This work is supported by the NIH awards R01-AI 143740 and R01-AI 146581 to JD

[1]Department of Computer Science, The Ohio State University, Columbus OH, USA

[2]GIG Statistical Consulting LLC, Columbus OH, USA.

[3]Department of Physics, The Ohio State University, Columbus OH, USA.

[4]Battelle Center for Mathematical Medicine, The Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus OH, USA

References

[1] J. Wu, W. Stewart, C. Jayaprakash, and J. Das, "Generalized method of moments improves parameter estimation in biochemical signaling models of time-stamped single-cell snapshot data," *bioRxiv*, 2022. [Online]. Available: https://www.biorxiv.org/content/early/2022/03/19/2022.03.17.484491