Integrating mathematical models with sparse time series data to predict disease dynamics

Project summary
Mathematical models are being increasingly used in medical research to understand and predict disease dynamics. For instance, ordinary differential equation (ODE) models are used to model how disease-indicating biomarkers, which can be measured from, e.g., patient’s blood samples, change over time. To have predictive abilities, such ODE models should ideally be fitted to patient-specific biomarker data that are collected at several time points. For practical reasons, however, medical time series data are often sparse and collected at time intervals that are irregular and vary between patients (Fig. 1). Currently, no best practices are in place for fitting mathematical models to such data. In this project, the PhD student will develop mathematical protocols for integrating ODE models with sparse and irregular biomarker time series data. These protocols will help researchers understand and predict biomarker dynamics and, by extension, disease dynamics.

This project will combine mathematical analysis, machine learning, and medical data to develop methods for understanding and predicting patient-specific disease dynamics.

Research questions
The PhD student will work towards answering Research questions 1-4.

Research question 1: How should we decide which (if any) ODE model to use for predicting biomarker and disease dynamics?

Research question 2: How often do patient-specific biomarker data need to be collected to parameterise predictive ODE models?

Research question 3: How should mathematical biomarker models and statistical methods be designed to handle inter-patient variability and biological stochasticity?

Research question 4: Can missing patient-specific biomarker data be patched by mathematical and computational methods that are informed by data from other patients?

Fig. 1: Ordinary differential equation models can be used to describe and predict patient-specific biomarker dynamics. Such models can be parameterised (or learned) from data that are measured from blood samples. ODE models that are learned from sparse data may have poor predictive abilities (patient 2, data). To model biomarker dynamics for patient 2, we therefore use data for patient 2 together with the learned model from patient 1.
Background

ODE models that are parameterised by patient-specific biomarker data have been used to predict the dynamics of e.g., Alzheimer’s disease[1], prostate cancer[2], and kidney diseases[3]. These ODE models have the general form

\[
d\frac{y(t, \theta)}{dt} = f(y(t, \theta)), \quad y(0, \theta) = y_0,
\]

where \(y(t, \theta)\) is the state vector that describes the system (the concentrations of dependent variables) at time \(t\) and \(y_0\) is the initial state vector. If \(f\) is assumed to be known, the parameter estimation problem entails learning the model parameters \(\theta\) from data. Methods from both frequentist statistics (e.g., the method of least squares) and Bayesian statistics (e.g., Hamiltonian Monte Carlo sampling) can be used to learn \(\theta\). In other cases, where \(f\) is not assumed to be known, machine learning methods based on Neural Networks (NNs) can be used to learn the full form of \(f(y(t, \theta))\) from data[4].

Both bottom-up approaches (formulating \(f\) and learning \(\theta\) from data) and top-down approaches (learning \(f\) from data) have merit in bio-mathematical modelling. The bottom-up approach enables us to directly integrate bio-medical knowledge into the form of \(f\), but the top-down approach allows for more general forms of \(f\). In this project, both bottom-up and top-down ODE learning methods will be studied, further developed, and evaluated against simulated and real data.

Aims and Methods

**Aim 1: Interrogate frequentist and Bayesian parameter estimation methods.** The student will compare methods from frequentist and Bayesian statistics that are currently being used to parameterise ODE models from time series biomarker data. These bottom-up methods are used to find \(\theta\) in Eq.1, and assume that \(f\) is known. The student will use evaluate the robustness of these methods when it comes to handling (i) variability between patients, (ii) irregular data, and (iii) biological stochasticity.

**Aim 2: Investigate ODE learning methods for modelling biomarker dynamics.** The student will use NN-methods to learn ODE models that describe biomarker dynamics. These top-down methods, which learn the full form \(f(y(t, \theta))\) in Eq.1, will be compared to those studied in Aim 1, and will be evaluated against simulated data. The NN-generated ODE models will also be evaluated against real data, and be compared to ODE models that are based on bio-medical theory and have previously been used to predict biomarker dynamics.

**Aim 3: Develop computational tools that enable parameter estimation for sparse and irregular time series data.** The student will develop computational tools that streamline the procedure of formulating and parameterising predictive ODE models from sparse and irregular time series data. The developed models and methods will first be trained and tested on simulated data, and will thereafter be evaluated against real data from clinical trials and biobanks.

Interdisciplinary significance

The project is interdisciplinary and designed to advance mathematical and bio-medical research.

**Significance for applied mathematics:** Integrating dynamic mathematical models with time series data is a central part of contemporary applied mathematics. In this project, mathematical and statistical methods for parameterising ODE models will be evaluated and further developed to be suitable for applications with sparse and irregular time series data in medicine and beyond. The methods developed in this project will be made accessible to other researchers via the distribution of open-access, non-proprietary computational tools.
Significance for medical research: Recent technical advances have resulted in a surge in medical data, including time series data for biomarkers that can be used to indicate disease progression and remission. The mathematical and statistical methods developed in this project will help researchers extract as much information as possible from already collected biomarker time series data. The methods will also suggest mathematically motivated protocols for the collection and reporting of future biomarker data, such that the data enable the parameterisation of predictive mathematical models.

Research team and environment

The successful candidate will work in the Division of Systems and Control at the Department of Information Technology. Advisor Dr Sara Hamis is an Assistant Professor in the Division of Systems and Control. Her research focuses on mathematical oncology, data-driven mechanistic mathematical modelling, and Bayesian statistics[5, 6]. Co-advisor Professor Stefan Engblom works in the Division of Scientific Computing and is an expert in scientific computing for bioscience applications[7]. Co-advisors Professor Tobias Sjöblom and Dr Jim Åkerrén Ögren at Uppsala University’s Department of Immunology, Genetics and Pathology have extensive knowledge in bio-medical, statistical, and practical aspects of using biomarker data to predict disease dynamics[8]. International collaborator Dr Morten Andersen is an Associate Professor at Roskilde University (Denmark) with expertise in modelling the dynamics of blood cancer and other diseases[3].

Prerequisites for applicants: Applicants should have a degree in mathematics, physics, statistics, machine learning, engineering, or a related field. Applicants should be scientifically curious problem-solvers who would enjoy working in interdisciplinary research teams. The ability to communicate in English is required. No previous bio-medical knowledge is required.

Financial support: This is a fully funded PhD position. The successful applicant will be funded by the Centre for Interdisciplinary Mathematics (50%) and the Division of Systems and Control (50%).

Contact: Prospective applicants are welcome to contact Sara Hamis with inquiries about the position (https://sarahamis.github.io/contact/).

References