Fast & accurate estimation of epidemiological parameters from pathogen genomes: with maximum likelihood and deep learning

Anna Zhukova

Institut Pasteur, Université de Paris, Paris, France

18 October 2022

Abstract

The interaction of epidemiological and evolutionary processes leaves a footprint in pathogen genomes. Phylodynamics leverages these footprints to estimate epidemiological parameters. Phylodynamic models estimate such parameters as the basic reproduction number, R_0, from topology and branch lengths of pathogen phylogenies (genealogies of pathogen sequences). In the era of constantly growing sequencing data, this approach can be especially useful for unravelling information on pathogen epidemics.

However, existing inference tools in a phylodynamic framework have not yet caught up with the sequencing speed. Computing time and numerical instability issues limit them to medium data sets ($\leqslant 500$ samples), while the accuracy of estimations should increase with data set size.

We solve this computational bottleneck in two ways: 1) via fine-tuned model-specific mathematical effort to improve likelihood computation for the Birth-Death Exposed-Infectious model, one of the most commonly used phylodynamic models, which targets pathogens featuring an incubation period (between the moment of infection and becoming infectious), such as Ebola and SARS-CoV2; and (2) via replacing a likelihood-based framework with a simulation-based (and therefore likelihood-free) deep learning approach for a larger spectrum of models. Our fast and accurate estimators are applicable to very large datasets ($> 10\,000$ samples).

About the presenter

Anna Zhukova is a senior bioinformatician at Institut Pasteur in Paris, France. Her research focuses on new phylogenetic methods, tools, and mathematical models to study pathogen transmission and evolution. For example, she is interested in such questions as: From where was Sars-CoV-2 introduced to France? Are there onward transmissions of drug-resistant HIV-1 among treatment-naive individuals? What is the expected number of cases directly infected by an Ebola-positive individual?

Anna's initial training was in software engineering and mathematics (St. Petersburg State University). She started working in bioinformatics during her PhD in Inria Bordeaux and could not stop since.