

Of pathogens and probabilities

DAVID SUDA,
MONIQUE BORG
INGUANEZ and
GIANLUCA URSINO

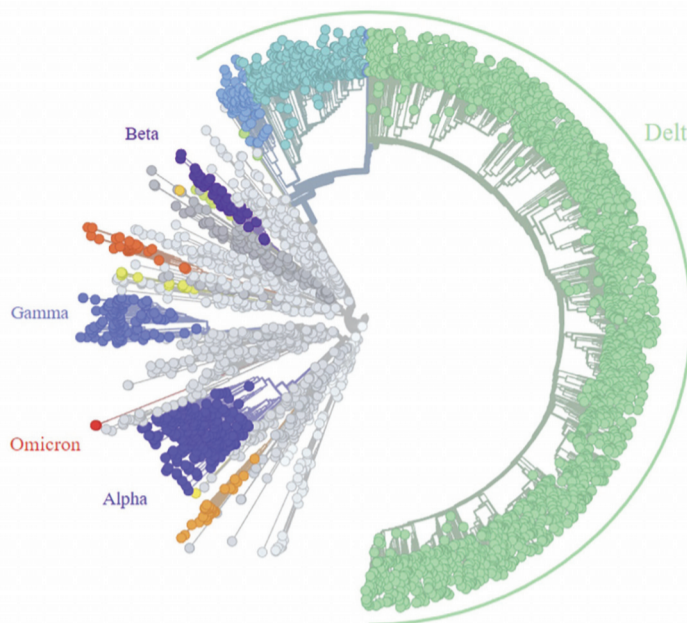
A brief overview shall be given of mathematical and statistical approaches which have been used to model pathogen dynamics - in particular, their transmission and their mutation. Both aspects have long been of interest to scientists, and with the advent of a pandemic in recent history, being able to effectively model these phenomena has become of utmost importance.

Epidemic models can be categorised as either deterministic or stochastic. Deterministic epidemic models are typically describing the mean behaviour of an epidemic, without taking into account any stochasticity. The simplest model of this type is the SIR model, where the terms S, I and R here refer to 'susceptible', 'infectious' and 'removed'. Variations and extensions of this model have been proposed with time. However, since the dynamics of communicable diseases is based on a collection of individual transmission dynamics, the applica-

bility of deterministic models is more suitable when dealing with larger populations, where randomness gets averaged out. On the other hand, stochastic (random) behaviour and also parameter uncertainty can be better incorporated within stochastic epidemic models. This would allow one to predict a distribution of possible health outcomes rather than a point forecast, and also allows one to provide confidence or credibility intervals. Stochastic epidemic models are typically developed within the conceptual framework of their deterministic counterparts.

When it comes to modelling the evolution of rapidly mutating pathogens such as viruses, one needs to look into an area called phylodynamic analysis. Phylodynamics, typically visualised via a phylodynamic tree, is related to phylogenetics, which is the study of the evolutionary history of a set of related species. However, phylodynamics also involves the estimation of epidemiological parameters, the root of a pandemic or epidemic, and also the divergence times of different species.

Phylogenetic and phylodynamic analysis is ideally nested



A phylogenetic tree of COVID-19 variants. PHOTO: WIKIMEDIA COMMONS: [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/MAIN_PAGE](https://commons.wikimedia.org/wiki/Main_Page)

within a Bayesian framework, as it makes sense to include some prior informative or non-informative knowledge on certain parameters. The Bayesian framework is ideal when one wants to incorporate parameter uncertainty within the model. The transmission model in Bayesian phylogenetic and phylodynamic analysis is usually either of the coalescent or birth-

death type. The substitution model, on the other hand, describes probabilistically the changes in the nucleotides of the viral genome sequences, which is modelled as a continuous-time Markov chain. Finally, the molecular clock model is assumed either to be strict, where the evolution rate is kept constant, or relaxed, where the evolution rate is not.

MYTH DEBUNKED

The frequentist approach to statistics is a special case of the Bayesian approach

MONIQUE BORG
INGUANEZ and
DAVID SUDA

The first step in any statistical analysis, be it frequentist or Bayesian, is the construction of a probabilistic model which describes, mathematically, the way the data under study has been generated. But the notion of probability differs in these two approaches to statistical inference.

The frequentist approach is based on the conventional notion of probability as frequency, commonly referred to as objective probability.

The (objective) probability of occurrence of an event, A, can be defined as the ratio of the number of ways in which A occurs compared with the total number of possible outcomes.

In contrast, in Bayesian statistics the notion of probability, commonly known as subjective probability, is one of a measure of one's degree of uncertainty about an event.

Under this approach, probability distributions are used to model knowledge about model parameters or hypotheses; this is done through the use of the famous Bayes' theorem.

Another key difference is that in the frequentist approach, results are based solely on the data collected; Bayesian inference, through the use of prior distributions, incorporates any information about the unknown parameters which is external from the data collected, such as results from previous similar studies. Nowadays, the common understanding is that some problems are better handled using frequentist methods while others with Bayesian methods.

PHOTO OF THE WEEK

```

GC AAAAGC AGGTCAATTTATATTCAGTATGAAAAGAATAAAAAGAACTACGGAA
GC AAAAGC AGGC AAACCATTTGAATGGATGTC AATCCGACTCTACTGTTCCCT
GC AAAAGC AGGTACTGATTCGAAATGGAAGATTTTGTGCGACAATGCTTCAA
GC AAAAGC AGGGGAAAATAAAAACAAACAAAATGAAAGCAAACTACTAGTT
GC AAAAGC AGGGGTTATACCATAGACACCAAGGCAAGCAATGGCCATCA
GC AAAAGC AGGGGATAATTCATTAACCATGAAGACTATCATTGCTTTGAGC
GC AAAAGC AGGGGAAAACAATGCTATCAATCAGGATTCCTGTTTCGCTCATAG
GC AAAAGC AGGGGTTCAATCTGTC AAAATGGAGAAAATAGTCTCTCTCTTG
GC AAAAGC AGGGGAAAATGATTCGAATCATTGTAATAGCGATACTGGCATCA
GC AAAAGC AGGGGATACGAAATGAACACTCAAAATCCTGGTATTCGCTCTGGT
GC AAAAGC AGGGGTCACCAATGGAGAAAATTCATCGCAATAGCAATGCTCTGG
GC AAAAGC AGGGGAAATTCACAAATTAACAAAATGGAAATAACAACTACTAAT
GC AAAAGC AGGGGTCACAAATGTACAAAATAGTACTAGTACTTGGCTCCTTG
GC AAAAGC AGGGGAAAATATCAAGAAAATCAAGATGAAGAAAGCAATCCCTCTT
GC AAAAGC AGGGGTCACAAATGGAAAATTCATCATTGAGCACTGTCTCTGG
GC AAAAGC AGGGGAAAATATTTAATGAAAACAGAAAATGGATATCAGAT
GC AAAAGC AGGGGAAAATGATTCGACTCATATTTGGTTGCACTGGCTCTGAGC
GC AAAAGC AGGGGAAAACAAAATGAACACTCAAAATCATTGTCATTCTAGTCTCT
GC AAAAGC AGGGGATATTTGTC AAAACAAAGAGAAATGGTGA TCAAAGTGCCTTA
GC AAAAGC AGGGTTAATAATCACTCACTCAGTGCACATCAAATCATGCCGTC

```

Before performing phylodynamic analysis, genome sequences obtained must be aligned at homologous positions. If genome sequences differ in length, gaps are inserted so that the aligned genome sequences will have the same length. These gaps can be interpreted as occurrences of other insertion or deletion mutations which occur when a nucleotide is added or deleted respectively in a genome sequence. There are a variety of sequence alignment tools that can be used including Clustal Omega, MAFFT, BLAST and MUSCLE. It is preferable that the gaps introduced are as few as possible as too many gaps can render the alignment meaningless.

SOUND BITES

- Change point analysis can be used to determine changes in the course of an epidemic, and thus also determine whether authorities' preventative measures to curtail the spread have been effective or whether there have been any behavioural changes in the population. In the 2020 paper *Time series analysis of COVID-19 infection curve: a changepoint perspective*, authors Feiyu Jiang, Zifeng Zhao and Xiaofeng Shao devise a changepoint model which determines changes in the exponential growth rate, and apply it to cumulative changes and cumulative deaths data of 30 countries in the initial stages of the pandemic.
- In the 2022 paper *Multiple Changepoint Analysis of COVID-19 Infection Progression and Related Deaths in the Small Island State of Malta*, the authors of this article apply an ensemble extension of the aforementioned changepoint model to similar COVID-19 epidemic data in Malta between March 7, 2020, and January 31, 2021, also postulating the potential causes for the inferred changepoints.

For more science news, listen to Radio Mocha on www.fb.com/RadioMochaMalta/.

DID YOU KNOW?

- The first epidemiological studies applying vital statistics date back to the 18th century in Britain. It was however towards the end of the 19th century that a new and more sophisticated statistical approach with a sound mathematical foundation emerged and eventually transformed the practice of epidemiology.
- Florence Nightingale, who is most popularly known as the pioneer of modern nursing, was also a pioneer in health statistics. One of her most significant contributions was the invention of polar area diagrams which depicted the causes of soldiers' deaths over two successive years in the Crimea.
- The first phylogenetic tree is credited to Charles Darwin.
- Gertrude Cox was an American statistician that earned the title of First Lady of Statistics, breaking the glass ceiling for other women to become statisticians.

For more trivia see: www.um.edu.mt/think.

CONTRIBUTORS

EDWARD DUCA
JEAN-PAUL EBEJER
DANIELLE MARTINE
FARRUGIA
JOSEF BORG
CLAUDIA BORG
ALEXANDER FARRUGIA

E-MAIL

SCI-SUNDAY@UM.EDU.MT