

GUIDING BAYESIAN MODEL CHOICE USING ABC,

WITH AN APPLICATION TO RAPIDLY EVOLVING INFECTIOUS DISEASES

Oliver Ratmann^{*}, Katia Koelle^{*}, Christophe Andrieu[†], Sylvia Richardson[‡]^{*} Biology Department, Duke University, USA[†] Dep of Mathematics, University of Bristol, United Kingdom[‡] Centre for Biostatistics, Imperial College London, United Kingdom

Model choice is a cardinal part of data analysis, and has been extensively discussed for scenarios in which the likelihood $f(x_0|\theta, M)$ of a scientific model M remains readily evaluable. In many areas of modern science, we now face increasingly often quantities of data for which $f(x_0|\theta, M)$ cannot be readily evaluated. Approximate Bayesian Computation (ABC) is a developing area of statistics that is motivated by the need to fit, assess and compare models that have a well-defined underlying generative process, but for which classical statistical procedures of estimation and goodness-of-fit encounter formidable computational challenges. ABC circumvents likelihood calculations by comparing model simulations with the data in terms of summary statistics.

To aid Bayesian model choice in this context, methods for model criticism and for model comparison have been introduced. We place these tools within the context of available Bayesian methodology, highlight their limitations and unique opportunities to Bayesian analysis, and propose new methods to guide model choice when it is only feasible to simulate from $f(\cdot|\theta, M)$

Model predictions have long been used to inform model choice; hence, ABC can be naturally embedded in a large array of available Bayesian methods. We explored the application of *sequential Bayesian tools* that comprise a training and testing phase, such as (approximate) posterior or partial predictive densities for model criticism, or posterior, pseudo or intrinsic Bayes' factors for model comparison when the likelihood cannot be evaluated. The extra volatility induced by simulating data sets typically offsets any gains in precision during the training phase, and we argue for using so-called simultaneous methods (1).

ABC enables us to check several aspects of the data, and thereby opens up multi-dimensional avenues towards more accurate model choice (2). We discuss in detail the power of multivariate predictive error densities for model criticism, as well as the accuracy of mutually constrained approximate Bayes' factors (aBF's) for model comparison. We also present aBF's for the purpose of multi-dimensional model criticism that simplify to the Savage-Dickey ratio and do not require the computation of the marginal likelihood.

The second part of the talk considers the application of ABC to the study of rapidly evolving infectious diseases such as influenza and HIV. Crucially, ABC makes possible to cross all the relevant biological scales involved (genetics, immunogenetics, ecology) when the summary statistics are suitably chosen. We will demonstrate and discuss the application of the aforementioned tools for model criticism and model choice to compare alternative hypotheses on the nature of immune escape and various forms of acquired immunity in influenza-A dynamics.

We conclude that, based on the above multi-dimensional techniques, we may check and compare complicated models even if aspects thereof remain unidentifiable. Moreover, these methods are particularly appealing in practice as they incur little or no extra computational cost compared to ABC.

References

- [1] Ratmann et al. Model criticism based on likelihood-free inference, with an application to protein network evolution. *PNAS USA*, 106(26):10576–10581, 2009.
- [2] Ratmann et al. Reply to Robert et al.: Model criticism informs model choice and model comparison. *PNAS USA*, 107(3):E6-7, 2010.