Malaria Transmission Modelling with Approximate Bayesian Computation

Ewan Cameron[1]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Department of Zoology, University of Oxford, Oxford, United Kingdom
dr.ewan.cameron@gmail.com

Abstract

Complex, biologically-inspired tranmission models for "micro-simulation"—the stochastic, computational realization of disease dynamics at the individual level within a mock human population—are emerging as a key tool in the field of malaria epidemiology. However, the long run times and high dimensionality of the parameter spaces for such models—ranging from ∼20 to 200 inputs plus a theoretically infinite-dimensional functional input profile of biting rate seasonality—pose substantial challenges for posterior inference via existing algorithms for Approximate Bayesian Computation (e.g. rejection ABC, SMC ABC, MCMC ABC). In this talk I will describe my ongoing work towards principled, yet efficient, posterior inference from these models via a fusion of ABC Indirect Inference and functional regression-based model emulation.

Keywords: Indirect inference; likelihood-free; functional regression; model emulation; micro-simulation.

1 Introduction

As a World Health Organisation "Collaborating Centre in Geospatial Disease Modelling" the Malaria Atlas Project[1] at the University of Oxford aims to produce accurate estimates of the global malaria burden, its geospatial distribution, and temporal trends. Crucial to our estimation framework is a robust model of the relationship between the parasite prevalence rate and the incidence of clinical disease within a community; for which, in the case of the *Plasmodium falciparum* parasite, both disease transmission theory and empirical observations indicate a complex, non-linear relationship and age-dependence owing to the effects of exposure-driven immunity [7], which in turn depends on a combination of the historical mean EIR (Entomological Innoculation Rate) and its seasonality profile [8].

Mathematical modelling of the disease transmission process within a population has a long history of success in the study of malaria (cf. the famous Ross-Macdonald models [9]), and is currently the focus of multiple ongoing efforts in the realm of burden estimation. State-of-the-art "micro-simulations" following mock populations of humans

[1] URL: http://www.map.ox.ac.uk

and multiple vector species allow for highly detailed simulations of the infection and treatment process, from which predictions of the prevalence–incidence relationship can be compiled. Although a vast body of past research into the mechanistics of the *Pf.* malaria parasite and the immunological responses to infection within individuals is available to inform some elements of these models, many key transmission parameters require further constraint against contemporary observational datasets [10]. Owing to the long computational run-times required to bring these "micro-simulations" to a steady state and extract mock observations, past studies have had to settle for maximum likelihood estimates under ad hoc approximations to the unknown likelihood function under the assumption of zero seasonality [11].

In this talk I will describe my work towards exploration of the full posterior of three contemporary "micro-simulation" models of malaria transmission via a fusion of functional regression-based model emulation and Approximate Bayesian Computation.

2 ABC-II with Functional Regression-based Model Emulation

For the purpose of ensemble forecasting we aim to average over predictions of the prevalence–incidence relationship from three popular contemporary codes for malaria "micro-simulation": the Griffin et al. (2014) model [6] ($\sim$20 free parameters), `OpenMalaria` [10] (up to $\sim$50 free parameters), and EMOD DTK [3] (up to $\sim$200 free parameters) constrained against a benchmark observational dataset compiled from a set of field studies chosen to ensure identifiability of at least the `OpenMalaria` parameters [11]. The Griffin et al. (2014) model has the fewest parameters and the fastest run time, owing in part to the existence of a corresponding steady-state solution that can be used for initialization, and is therefore the focus of our methodological validation. However, the EMOD DTK has never before been fitted to observations and is thus the key scientific target of our analysis.

We build upon previous efforts to fit the `OpenMalaria` model in which maximum likelihood fitting was conducted under an ad hoc approximation to the true likelihood function under the assumption of no seasonality by first connecting the full posterior inference challenge to the framework of ABC Indirect Inference (II) [5, 2]. We then show how the parameters of the II model can be estimated to adequate accuracy, and at much greater computational efficiency, using far smaller mock populations than the original analysis; and thus how a large suite of small population simulations can be used to train a functional regression-based [4, 1] model emulator able to provide an effective proxy for use in ABC-II.

The age-structured prevalence-incidence relationships so derived have been combined with our latest geospatial prevalence models to produce the most up-to-date estimate of the clincal malaria burden across the African continent.

References

[1] Ciollaro, M., Cisewski, J., Freeman, P., Genovese, C., Lei, J., O'Connell, R., Wasserman, L. (2014). "Functional Regression for Quasar Spectra." *preprint.*

[2] Drovandi, C., Pettitt, A.N. (2013). "Bayesian Indirect Inference." *preprint.*

[3] Eckhoff, P.A. (2011). "A malaria transmission-directed model of mosquito life cycle and ecology." *Malaria Journal*, **10**, 303.

[4] Ferraty, F., Vieu, P. (2004). "Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination." *Journal of Nonparametric Statistics*, **16**, 111.

[5] Gleim, A., Pigorsch, C. (2013). "Approximate Bayesian Computation with Indirect Summary Statistics." *preprint*.

[6] Griffin, J.T., et al. (2014). "Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa." *Nature Communications*, **5**, 3136.

[7] Patil, A.P., et al. (2009). "Defining the relationship between Plasmodium falciparum parasite rate and clinical disease: statistical models for disease burden estimation." *Malaria Journal*, **8**, 186.

[8] Roca-Feltrer, A., et al. (2010). "The age patterns of severe malaria syndromes in sub-Saharan Africa across a range of transmission intensities and seasonality settings." *Malaria Journal*, **9**, 282.

[9] Smith, D.L., et al. (2012). "Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens." *PLOS Pathogens*, **8**(4), e1002588.

[10] Smith, T., et al. (2006). "Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of Plasmodium falciparum malaria: Overview." *Am. J. Trop. Med. Hyg.*, **75**(2), supp. 1-10.

[11] Smith, T., et al. (2012). "Ensemble modeling of the likely public health impact of a pre-erythrocytic malaria vaccine." *PLOS Medicine*, **9**(1), e1001157.