Methods for efficient statistical inference of stochastic dynamical systems

Aaron A. King

University of Michigan, United States, kingaa@umich.edu

Keywords: partially observed Markov process, stochastic dynamical system, likelihood, plug-and-play inference

Many important questions in the epidemiology and ecology of communicable diseases are naturally formulated as questions about the structure and parameterization of dynamical systems models. Retrospective analyses of data on recurrent outbreaks of such infections can yield answers to these questions, provided one can construct a tight connection between a stochastic model and data. Because of the centrality of questions regarding model structure, however, and due to the general intractability of most models of interest, there is great value in methods that place minimal restrictions on the form of models and allow comparison of the explanatory and predictive power of disparate models on an efficient and objective basis. In this talk, I first give examples of some questions of this nature drawn from the epidemiology of pertussis, cholera, and polio [4, 6, 7]. Key issues involve route or mode of transmission, the nature, strength, and durability of infection- and vaccine-induced immunity, and the ecological mechanisms promoting persistence and resisting eradication of the pathogen.

In the next part of the talk, I motivate and explain the concept of "plug-and-play" inference algorithms and highlight three particularly useful algorithms from this class: (1) particle Markov chain Monte Carlo [1], (2) synthetic likelihood [9, 2], and (3) iterated filtering [3]. For each of these methods, I explain its key principles, sketch the algorithm structure, and describe its advantages and disadvantages in terms of statistical and computational efficiency. These methods compare favorably to the "standard" and time-honored approaches of data-augmentation Markov chain Monte Carlo and the various Kalman filter extensions and have appreciable advantages over newer methods such as approximate Bayesian computation [8]. On the other hand, they can be computationally demanding and face considerable challenges in dealing with high-dimensional data. I argue that, on balance, there is much to be gained from the capacity to effectively explore the complex, high-dimensional, nonlinear likelihood surfaces associated with questions of scientific interest. Doing so is greatly facilitated by the availability of software implementing these, and other, plug-and-play algorithms within a flexible modeling environment [5].

To conclude the talk, I demonstrate iterated filtering in practice as a method for answering some of the questions raised in the first part of the talk. In particular, I show how the likelihood profile and associated tools can be used not only to quantify statistical support for scientific findings but to shed light on how nonlinearity and stochasticity interact to shape infectious disease dynamics.

Bibliography

[1] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 72(3):269–342, 2010.

[2] Matteo Fasiolo, Natalya Pya, and Simon Wood. Statistical inference for highly non-linear dynamical models in ecology and epidemiology. arXiv:1411.4564, 2014.

[3] Edward L. Ionides, Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A. King. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences of the U.S.A.*, 112(3):719–724, January 2015.

[4] Aaron A. King, Edward L. Ionides, Mercedes Pascual, and Menno J. Bouma. Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880, August 2008.

[5] Aaron A. King, Dao Nguyen, and Edward L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, in press. arXiv:1509.00503

[6] F. M. G. Magpantay, M. Domenech de Cellès, P. Rohani, and A. A. King. Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity. *Parasitology*, in press.

[7] Micaela Martinez-Bakker, Aaron A. King, and Pejman Rohani. Unraveling the transmission ecology of polio. *PLoS Biology*, 13(6):e1002172, June 2015.

[8] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society, Interface*, 6(31):187–202, Feb 2009.

[9] Simon N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1104, August 2010.