Statistical Bioinformatics and Stochastic Systems Biology
Newcastle May 2009

Approximate Bayesian Computation under model uncertainty, with application to protein network evolution

Sylvia Richardson

Imperial College London, UK
sylvia.richardson@imperial.ac.uk

Joint work with **Oliver Ratmann** (Imperial College) and **Christophe Andrieu** (University of Bristol).

In many areas of computational biology, the likelihood $f(x_0 \mid \theta, M)$ of a scientific model M is intractable, typically because interesting models are highly complex. This hampers scientific progress in terms of iterative data acquisition, parameter inference, model checking and model refinement within a Bayesian framework. Nevertheless, given a value of θ, it is typically possible to simulate data from $f(\cdot \mid \theta, M)$. ABC proposes to infer θ by comparing simulated data x to the observed data x_0, in terms of a (real-valued) univariate discrepancy ρ that combines a set of (computationally tractable) summaries $\mathbb{S} = (S_1, \ldots, S_k, \ldots, S_K)$ (1). In its simplest form, values of θ for which the discrepancies are within $\tau \geq 0$ are retained to define the approximate likelihood.

After briefly reviewing current developments, I will discuss how to extend the current framework to consider *simultaneously* model inference and model criticism, reinterpreting a set of several discrepancies: $\{\rho_k\big(S_k(x), S_k(x_0)\big)\}$ as *realizations* of real-valued error terms, denoted by $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_K)$. A theoretical framework, *ABC under model uncertainty*, will be presented as well as a description of possible MCMC algorithms for inferring the joint approximate posterior distribution of summary errors and model parameters (2). By making probabilistic statements of mismatch between the model and the data, model criticism is facilitated and guidance on how to improve the models is gained. Finally, the benefit of incorporating model diagnostics within the ABC framework will be demonstrated by contrasting three qualitative models of protein network evolution to analyse the protein interaction datasets of *Helicobacter pylori* and *Treponema pallidum* (3).

(1) P. Marjoram and S. Tavaré. Modern computational approaches for analysing molecular genetic variation data. *Nat Rev Genet*, 7(10):759–770, 2006.

(2) O. Ratmann, C. Andrieu, C. Wiuf, and S. Richardson. Model criticism with likelihood-free inference, with an example from evolutionary systems biology. *Under review*, 2009.

(3) O. Ratmann, O. Jørgensen, T. Hinkley, M. P.H. Stumpf, S. Richardson, and C. Wiuf. Using likelihood-free inference to compare evolutionary dynamics of the protein networks of H.pylori and P.falciparum. *PLoS Computational Biology*, 3 (11): e230, 2007.