Applications of nested sampling in systems biology

Stuart Aitken

Centre for Systems Biology at Edinburgh, University of Edinburgh, UK

http://www.csbe.ed.ac.uk

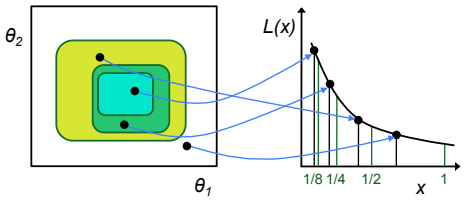
Stochastic models are commonly used in systems biology to represent the interaction of small numbers of molecules, and the discrete states that a molecule might adopt. The optimisation of complex stochastic models is challenging as, typically, they cannot be solved analytically.

Nested sampling is an effective method for sampling the posterior distributions of model parameters. The samples are obtained as a by-product of calculating the *Bayesian evidence*. Nested sampling requires a likelihood function, and, in the context of systems biology, the extent to which the data is explained by a given set of model parameters can be computed by an approximate log likelihood function derived from a number of Gillespie simulations. This optimisation strategy is therefore generic, and applicable to kinetic data and steady-state distributions.

We have demonstrated that this approach performs well as an optimiser for a number of systems biology models, including models of circadian rhythms. The method can also be used for model comparison – which will be the topic of future work.

Nested sampling

Nested sampling explores the Bayesian evidence, transforming the multi-dimensional integral for the evidence into a one-dimensional integral over the prior mass (Skilling, 2006). The sorted likelihood function $L(x)$ is used as an evolving constraint in the generation of a set of objects $\{x_0..x_i\}$, each object is an array of values randomly sampled from the prior range of a model parameter.



Let $x(L)$ be the prior mass enclosed within the contour $L(\theta) = L$, § and $L(x)$ be the contour value such that the volume enclosed is x.

$Z = \int dx\, L(x)$

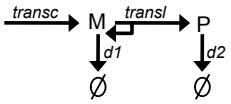
Samples ● are drawn uniformly from $L(x) > L^*$

Given a set of n *active* objects, the worst is replaced by a new object x_{new}, subject to the constraint $L(x_{new}) > L^*$. The new object is discovered by an exploration method that takes one of the remaining $n-1$ objects as a starting point. The constraint L^* is then set equal to the log likelihood of the worst object, and the process repeats. The objects eliminated from the active set are added to the set of *posterior* samples, and can be analysed further. The new object returned by the exploration method must be an unbiased sample of the prior subject to the constraint on the likelihood L^*: unbiased sampling is necessary for correctly computing Z, and can be achieved by the correct choice of exploration function (Murray, 2007).

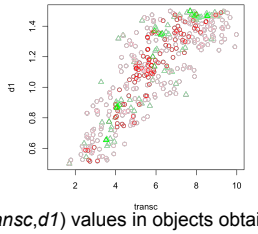
Nested sampling requires a likelihood function to be defined in order to compute the probability of the data given the model. An approximate likelihood can be derived from a number of realisations of a model obtained by stochastic simulation.

§ following MacKay http://www.inference.phy.cam.ac.uk/bayesys/

A simple model of transcription and translation

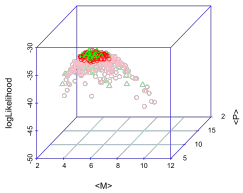


A stochastic model of the transcription of mRNA (M) and the translation of protein (P) from mRNA was fitted to a synthetic data set of samples from the distributions of M and P. Nested sampling reveals the obvious trade-off between *transc* and *d1* in establishing the mean <M>.



Circles are (*transc*,*d1*) values in objects obtained when n=100; triangles are values obtained when n=20. Lighter colours indicate posterior samples, darker colours indicate active objects.

The log likelihood is maximised when <M> = *transc/d1* = 5.14 and <P> = *transc*×*tranls/d1*×*d2* = 10.28. The posterior samples generated by nested sampling converge on these ratios as shown below.



Future work will consider more complex systems biology models, and the use of Bayesian evidence for model selection.

References

Murray, I. (2007) *Advances in Markov chain Monte Carlo methods*. PhD Thesis University College London.

Skilling, J. (2006) Nested sampling for Bayesian computations. *Proc. ISBA* 2006.

SA is supported by a Wellcome Trust Value In People award.

THE UNIVERSITY of EDINBURGH

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

"Pioneering research and infrastructure development to model the dynamic aspects of biology"

BBSRC bioscience for the future

EPSRC Engineering and Physical Sciences Research Council