Dizzy-Beats

Citation for published version:

Digital Object Identifier (DOI):
10.1093/bioinformatics/btv062

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Bioinformatics

Publisher Rights Statement:
© The Author(s) 2015. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Dizzy-Beats: a Bayesian Evidence analysis tool for systems biology

Stuart Aitken 1,∗Alastair M. Kilpatrick 2,3 and Ozgur E. Akman 4

1 MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK
2 School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK
3 Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA
4 Centre for Systems, Dynamics and Control, College of Engineering, Mathematics & Physical Sciences, University of Exeter EX4 4QF UK

ABSTRACT

Motivation: Model selection and parameter inference are complex problems of long-standing interest in systems biology. Selecting between competing models arises commonly as underlying biochemical mechanisms are often not fully known, hence alternative models must be considered. Parameter inference yields important information on the extent to which the data and the model constrain parameter values.

Results: We report Dizzy-Beats, a graphical Java Bayesian evidence analysis tool implementing nested sampling - an algorithm yielding an estimate of the log of the Bayesian evidence \(Z\) and the moments of model parameters, thus addressing two outstanding challenges in systems modelling. A likelihood function based on the \(L_1\)-norm is adopted as it is generally applicable to replicated time series data.

Availability: http://sourceforge.net/p/bayesevidence/home/Home/

Contact: s.aitken@ed.ac.uk

1 INTRODUCTION

Bayesian methods provide a sound basis for ranking alternative systems biology models and for characterising the extent to which parameters are constrained by models and data (Kirk et al., 2013). Markov Chain Monte Carlo (MCMC) methods have been applied to model selection (Schmidl et al., 2012) and to parameter inference in systems biology (Hug et al., 2013; Kanodia et al., 2014), but often require considerable algorithmic and conceptual development. Nested sampling promises to ease these complex computational tasks: Recent biological applications include (Aitken and Akman, 2013; Kirk et al., 2013; Pullen and Morris, 2014).

General purpose code for nested sampling is available in R (Skilling, 2006; Aitken and Akman, 2013), and biological applications of the MultiNest tool (Feroz et al., 2013) have been reported (Kirk et al., 2013; Pullen and Morris, 2014). A C-based command-line application implementing nested sampling and providing an SBML interface has recently been released (Johnson et al., 2014), but no graphical tool is currently available. Thus we sought to add nested sampling to the widely-used Dizzy chemical kinetics simulation tool (Ramsey et al., 2005) (over 200 citations as of November 2014). While doing so we also added an optimisation function and SBML 3.1 compatibility. However, as Dizzy’s command language has operators that cannot be captured in SBML 3.1, and SBML 3.1 has features not supported by Dizzy, this feature is restricted to the intersection of the modelling languages.

2 METHODS

Nested sampling calculates two of the central results of Bayesian inference: the posterior distribution \(P(\theta | D, H_i)\) of the parameters \(\theta\), and the evidence \(P(D | H_i)\), that is, the support for the data \(D\) under hypothesis \(H_i\) (Skilling, 2006), through a sampling strategy. A selection between two alternative models \(H_0\) and \(H_1\) can be made by calculating the ratio of their posterior probabilities (1), a calculation that can be decomposed into the Bayesian evidence \((Z_0\) and \(Z_1\) ) and the prior probability of the respective hypotheses.

\[
P(H_1 | D) \quad P(H_0 | D) = \frac{P(D | H_1) P(H_1)}{P(D | H_0) P(H_0)} = \frac{Z_1 P(H_1)}{Z_0 P(H_0)}
\]

The evidence (2) is a scalar quantity that can be viewed as an integral of the likelihood \((L)\) over the elements of mass \((dX = \pi(\theta)d\theta)\) associated with the prior density \(\pi(\theta)\).

\[
Z = \int L(\theta) \pi(\theta) \, d\theta
\]

The likelihood \((L)\) over the elements of mass \((dX = \pi(\theta)d\theta)\) associated with the prior density \(\pi(\theta)\). The prior mass can be accumulated from its elements \((dX)\) in any order. The enclosed prior of likelihood \(> \lambda\) can be defined (3), and this allows the evidence to be written as a one-dimensional integral of the (inverse) likelihood \(L(X)\) over the unit range taking the enclosed prior mass \(X\) to be the primary variable (4) (Skilling, 2006).

\[
X(\lambda) = \int_{L(\theta) > \lambda} \pi(\theta) \, d\theta
\]

\[
Z = \int_{0}^{1} L(X) \, dX
\]

\[
L(X(\lambda)) = \lambda
\]

Given a sequence of decreasing values \(0 < X_m < \ldots < X_2 < X_1 < 1\) where the likelihood \(L_i = L(X_i)\) can be evaluated, the evidence can be approximated numerically as a weighted sum. Inferences about the posterior can be obtained from the sequence of \(m\) discarded points generated by sampling, \(P\). Each point is assigned the weight \(w_i = L(\theta_i)w_i / Z\), from which the first and second moments of each parameter in \(\theta\) can be estimated – for more details see Skilling (2006) and Aitken and Akman (2013).

The size of the population of active points (points \(\theta_i\) within the evolving constraint \(L(\theta_i) > \lambda\) used to sample the parameter space is the only
3 DISCUSSION

Dizzy-Beats is a graphical application for simulating and optimising systems models based on an established simulator (Ramsey et al., 2005) and its simple textual model syntax, to which we have added SBML 3.1 import/export functionality. Uniquely, Dizzy-Beats provides model comparison and parameter inference functions through the nested sampling algorithm in a graphical application. Comparable functions are implemented in BioBayes (Vyshemirsky and Girolami, 2006), however, users must edit the XML representation of the model should they wish to make modifications. SYSBIONS (Johnson et al., 2014) implements nested sampling but all interaction is via the command-line. The use of a likelihood based on the L1-norm derived from biological replicate data makes fewer assumptions than a Gaussian error model (Vyshemirsky and Girolami, 2008; Johnson et al., 2014), and is less computationally complex than a transitional likelihood function derived from reaction propensities (Heron et al., 2007; Aitken and Akman, 2013).

ACKNOWLEDGEMENT

Funding: This work was funded by BB/E023461/1 (Bayesian evidence analysis tools for systems biology; S.A. and O.E.A.).

REFERENCES


