Cell motility models and inference for dynamic systems

or

Plug-and-play inference: What? Why? When? How?

JSM2012

Discussion of "Bayesian Spatio-Dynamic Modeling in Cell Motility Studies," by Manolopoulou et al.

August 1, 2012

Edward Ionides

The University of Michigan, Department of Statistics

Motivating plug-and-play inference for cell motility

- Multiple models have been proposed for cell movements.
- If the choice of model matters, one wants to compare previous suggestions and investigate improvements.
- Wanted: an inference framework for general classes of dynamic models.
- Simulation-based methods offer generality. They can carry out likelihood-based, feature-based or Bayesian inference.
- Modern simulation-based inference methods do not generate long trajectories, but make cunning use of many carefully selected short simulations. Such methods are said to be **Plug-and-play**.

An analogy with deterministic models

- Scientists and engineers working with ordinary differential equation (ODE) models often expect to carry out investigations using numerical methods.
- Practitioners using ODE models do not expect to be limited to models with convenient analytic properties.
- Plug-and-play methodology gives the same modeling freedom to practitioners using stochastic dynamic models.

Specifying a Markov model $X(t)$ and observable process $Y(t)$

`rprocess(x0,t0,t1,theta)` : a draw from the transition model, $X(\texttt{t1})$ given $X(\texttt{t0}) = \texttt{x0}$ with parameter vector `theta`.

`dprocess(x0,x1,t0,t1,theta)` : the transition density at `x1`.

`rmeasure(x1,t1,theta)` : a draw from the measurement model, $Y(\texttt{t1})$ given $X(\texttt{t1}) = \texttt{x1}$ with parameter vector `theta`.

`dmeasure(y1,x1,t1,theta)` : the measurement density at `y1`.

These `process` and `measure` functions define a **state space model**.

An algorithm calling `rprocess` but not `dprocess` is **plug-and-play**.

Software representing models via specification of `rprocess` and/or `dprocess` and/or `rmeasure` and/or `dmeasure`

- the R package **`pomp`** implements various proposed methodologies for partially observed Markov process (POMP) models, also known as state space models.

- If you have simulation code for a dynamic model, and you add a front end matching the required format of **`rprocess`**, then you are ready to do inference. Of course, you also need data and a measurement model.

- Once you have ‘plugged’ the simulator into **`pomp`**, you can ‘play’ with parameter estimation and hypothesis testing.

MCMC and EM are not plug-and-play

- Markov chain Monte Carlo (MCMC) and Expectation-Maximization (EM) algorithms for POMP models require **`dprocess`**. Therefore, they are not plug-and-play.

- MCMC and EM algorithms have another problem for POMP models. For both these data augmentation methods, convergence difficulties arise when measurements on the unobserved state process would be highly informative about parameters.

- Writing general software for MCMC and EM algorithms applicable to POMP models is extremely difficult, even if **`dprocess`** is available. Standard computational frameworks such as BUGS are not usually successful on this class of models.

Plug-and-play methods implemented in `pomp`

	Frequentist	Bayesian
Full-information	iterated filtering	particle MCMC
Feature-based	simulated moments	ABC

- Many excitement-generating recent methods are plug-and-play.
- particle MCMC (Andrieu et al, 2010, *JRSSB*) and approximate Bayesian computation (ABC; Sisson et al, *PNAS*) are hot Bayesian methods.
- A simulated feature comparison method was recently proposed (Wood 2010, *Nature*).
- Iterated filtering using sequential Monte Carlo enables likelihood-based plug-and-play inference (Ionides et al, 2006, *PNAS*).

Plug-and-play in other settings

- **Optimization**. Methods requiring only evaluation of the objective function to be optimized are sometimes called **gradient-free**. This is the same concept as plug-and-play: the code to evaluate the objective function can be *plugged into* the optimizer.

- **Complex systems**. Methods to study the behavior of large simulation models that only employ the underlying code as a “black box” to generate simulations have been called **equation-free**.

  ◊ This is the same concept as plug-and-play, but we prefer our label!

  ◊ A typical goal is to determine the relationship between macroscopic phenomena (e.g. phase transitions) and microscopic properties (e.g. molecular interactions).

The cost of plug-and-play

- Analytic tractability should save computational time.

  ◊ For optimization, one expects that analytic tractability of derivatives will improve computational performance.

  ◊ Nevertheless, gradient-free optimization is a frequently used tool.

- For huge models, plug-and-play methods are not yet practical. Unless one can avoid the curse of dimensionality, they may never be practical.

- For models and datasets comparable to Manolopoulou et al, plug-and-play methods have been successfully demonstrated (Bhadra et al, 2011, *JASA*; King et al, 2008, *Nature*).

Statistics: Growing to serve a data-dependent society

- Plug-and-play is a property of **algorithms**.
  - ◊ for example, EM and iterated filtering both carry out maximum likelihood estimation, but only the second is plug-and-play.
  - ◊ for complex data, properties of algorithms can be as important as properties of the resulting estimators.
- The plug-and-play property is very useful for scientific investigations seeking quantitative explanations of data from dynamic systems.

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 72:269–342.

Bhadra, A., Ionides, E. L., Laneri, K., Pascual, M., Bouma, M., and Dhiman, R. C. (2011). Malaria in Northwest India: Data analysis via partially observed stochastic differential equation models driven by Lévy noise. *Journal of the American Statistical Association*, 106:440–451.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the USA*, 103:18438–18443.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454:877–880.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the USA*, 104(6):1760–1765.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1104.