POMP inference via iterated filtering

Edward Ionides
University of Michigan, Department of Statistics

Lecture 3 at Wharton Statistics Department
Thursday 27th April, 2017

Slides are online at
`http://dept.stat.lsa.umich.edu/~ionides/talks/upenn`

From filtering to inference

- The POMP prediction and filtering recursions provide the likelihood for a given parameter θ.
- Successful filtering does not directly enable likelihood-based (frequentist or Bayesian) inference.
- Subtle numerical issues arise in briding this gap.
- Two "obvious" approaches to parameter inference via SMC fail in all but simple problems:
  - Apply a black-box optimizer such as Nelder-Mead to the SMC evaluation of the likelihood.
  - Carry out Bayesian inference by SMC with θ added to the POMP as a static parameter.
- Standard MCMC and EM algorithms also work poorly on POMP models.

SMC with perturbed parameters

- One natural way to coerce a filter into providing inference on parameters is to augment the latent variable with the parameter vector.
- We can change the latent variable space to $\breve{\mathbb{X}} = \mathbb{X} \times \Theta$ and the latent variable to $\breve{X}(t) = \big(X(t), \theta\big)$.
- However, θ has trivial dynamics: it is constant!
- This leads to poor numerical stability of filters.
- Filter stability is restored by adding stochastic dynamics, $\theta(t)$
- This idea was used heuristically by Kitagawa (1998) and Liu and West (2001). These methods are numerically stable, but they don't well approximate the goal of inference for fixed parameters.
- Better methods were needed.

An iterated filtering algorithm: IF1

- Iterated filtering algorithms adapt sequential Monte Carlo (SMC) into a tool for inference on θ by repeated the filtering process with a time-varying $\theta(t)$ having decreasing stochastic perturbations.
- We call IF1 the iterated filtering algorithm of Ionides, Bretó and King (2006). IF1 uses SMC filtering with perturbed parameters to approximate the derivative of the log likelihood.

"Powerful new inferential fitting methods (Ionides, Bretó and King, 2006) considerably increase the accuracy of outbreak predictions while also allowing models whose structure reflects different underlying assumptions to be compared. These approaches move well beyond time series and statistical regression analyses as they include mechanistic details as mathematical functions that define rates of loss of immunity and the response of vector abundance to climate."

Dobson, A. (2014). Mathematical models for emerging disease. *Science*.

An improved iterated filtering algorithm: IF2

- IF2 has an entirely different theoretical foundation from IF1.
- IF2 also happens to perform considerably better, though the theory we see below will not explain why!
- IF2 is based on the observation that iterating a Bayes map **(repeated application of Bayes theorem on the same data, using the posterior from one application of Bayes theorem as the prior to the next)** converges to an MLE.
- We can add perturbations to the parameters in each iteration, to stabilize the numerical approximation, without losing this convergence.

Input and output for the IF2 algorithm

input:

Simulator for latent process initial density, $f_{X_0}(x_0\,;\theta)$

Simulator for transition density, $f_{X_n|X_{n-1}}(x_n\,|\,x_{n-1}\,;\theta)$, n in $1\!:\!N$

Evaluator for measurement density, $f_{Y_n|X_n}(y_n\,|\,x_n\,;\theta)$, n in $1\!:\!N$

Data, $y^*_{1:N}$

Number of iterations, M

Number of particles, J

Initial parameter swarm, $\{\Theta^0_j,\ j$ in $1\!:\!J\}$

Perturbation density, $h_n(\theta\,|\,\varphi\,;\sigma)$, n in $1\!:\!N$

Perturbation sequence, $\sigma_{1:M}$

output: Final parameter swarm, $\{\Theta^M_j,\ j$ in $1\!:\!J\}$

Algorithms that specify the dynamic model via a simulator are said to be **plug-and-play**. This property ensures applicability to the broad class of models for which a simulator is available.

IF2: iterated SMC with perturbed parameters

```
For $m$ in $1\!:\!M$ [$M$ filtering iterations, with decreasing $\sigma_m$]
    $\Theta^{F,m}_{0,j} \sim h_0(\,\cdot \mid \Theta^{m-1}_j\,; \sigma_m)$ for $j$ in $1\!:\!J$
    $X^{F,m}_{0,j} \sim f_{X_0}(x_0; \Theta^{F,m}_{0,j})$ for $j$ in $1\!:\!J$
    For $n$ in $1\!:\!N$ [SMC with $J$ particles]
        $\Theta^{P,m}_{n,j} \sim h_n(\,\cdot \mid \Theta^{F,m}_{n-1,j}, \sigma_m)$ for $j$ in $1\!:\!J$
        $X^{P,m}_{n,j} \sim f_{X_n|X_{n-1}}(x_n \mid X^{F,m}_{n-1,j}; \Theta^{P,m}_j)$ for $j$ in $1\!:\!J$
        $w^m_{n,j} = f_{Y_n|X_n}(y^*_n \mid X^{P,m}_{n,j}; \Theta^{P,m}_{n,j})$ for $j$ in $1\!:\!J$
        Draw $k_{1:J}$ with $\mathbb{P}(k_j = i) = w^m_{n,i} \Big/ \sum_{u=1}^{J} w^m_{n,u}$
        $\Theta^{F,m}_{n,j} = \Theta^{P,m}_{n,k_j}$ and $X^{F,m}_{n,j} = X^{P,m}_{n,k_j}$ for $j$ in $1\!:\!J$
    End For
    Set $\Theta^m_j = \Theta^{F,m}_{N,j}$ for $j$ in $1\!:\!J$
End For
```

IF2: iterated SMC with perturbed parameters

```
For $m$ in $1\!:\!M$
    $\Theta^{F,m}_{0,j} \sim h_0(\,\cdot \mid \Theta^{m-1}_j \,; \sigma_m)$ for $j$ in $1\!:\!J$
    $X^{F,m}_{0,j} \sim f_{X_0}(x_0; \Theta^{F,m}_{0,j})$ for $j$ in $1\!:\!J$
    [carry out SMC on an extended model, with the time-varying parameters included in the latent state, initialized at $(X^{F,m}_{0,j}, \Theta^{F,m}_{0,j})$]
    Set $\Theta^m_j = \Theta^{F,m}_{N,j}$ for $j$ in $1\!:\!J$
End For
```

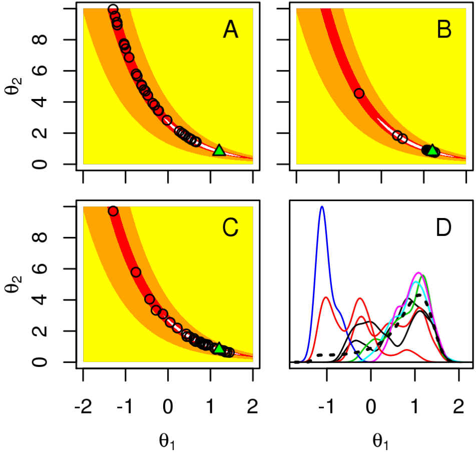
Numerical examples

- We compare IF1, IF2 and the particle Markov chain Monte Carlo (PMCMC) method of Andrieu et al (2010).
- PMCMC is an SMC-based plug-and-play algorithm for full-information Bayesian inference on POMPs.
- Computations were done using the `pomp` R package:

  King, Nguyen & Ionides (2016). "Statistical inference for partially observed Markov processes via the R package `pomp`." *Journal of Statistical Software*.

- Data and code reproducing our results are a supplement to

  Ionides, Nguyen, Atchadé, Stoev & King (2015). "Inference for dynamic and latent variable models via iterated, perturbed Bayes maps." *Proceedings of the National Academy of Sciences of the USA*.



Toy example.

$X(t) = \left(\exp\{\theta_1\}, \theta_2 \exp\{\theta_1\}\right)$,
constant for all t.

100 independent observations:
Given $X(t) = x$,

$$Y_n \sim \text{Normal}\left[x, \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

A. IF1 point estimates from 30 replications and the MLE (green triangle).
B. IF2 point estimates from 30 replications and the MLE (green triangle).
C. Final parameter value of 30 PMCMC chains with 10^4 filtering iterations.
D. Kernel density estimates from 8 of these 30 PMCMC chains, and the true posterior distribution (dotted black line).

Why is IF2 so much better than IF1 on this problem?

- IF1 updates parameters by a linear combination of filtered parameter estimates for the extended model with time-varying parameters.
- Taking linear combinations can knock the optimizer off nonlinear ridges of the likelihood function.
- IF2 does not have this vulnerability.
- A heuristic argument suggests that IF2 has 2nd order convergence (as if the 1st and 2nd derivatives were computable) whereas IF1 has 1st order convergence. It is an open problem to formalize that.

Epidemiological applications: A review of disease dynamics

- Communicable diseases have long had major global health impact (malaria, tuberculosis, measles, etc).
- Emerging and re-emerging diseases need to be understood and controlled (HIV, Ebola, bird flu, SARS, Zika, dengue, etc).
- Central to math models is an infected population, $I(t)$, which interacts with a susceptible population, $S(t)$. Susceptible individuals become infected at a nonlinear rate $\beta\, I(t)\, S(t)$, where β is a contact rate.
- The inherent stochasticity of biological populations, and our partial ability to observe epidemics, therefore lead to nonlinear POMP model inference problems.

Application to a cholera model

The study population $P(t)$ is split into susceptibles, $S(t)$, infecteds, $I(t)$, and k recovered classes $R_1(t), \ldots, R_k(t)$. The state process $X(t) = (S(t), I(t), R_1(t), \ldots, R_k(t))$ follows a stochastic differential equation driven by a Brownian motion $\{B(t)\}$,

$$\begin{aligned}
dS &= \Big\{k\epsilon R_k + \delta(P - S) - \lambda(t)\, S\Big\}dt + dP - (\sigma I/P)\, dB, \\
dI &= \Big\{\lambda(t)\, S - (m + \delta + \gamma)I\Big\}dt + (\sigma I/P)\, dB, \\
dR_1 &= \Big\{\gamma I - (k\epsilon + \delta)R_1\Big\}dt, \\
&\;\vdots \\
dR_k &= \Big\{k\epsilon R_{k-1} - (k\epsilon + \delta)R_k\Big\}dt.
\end{aligned}$$

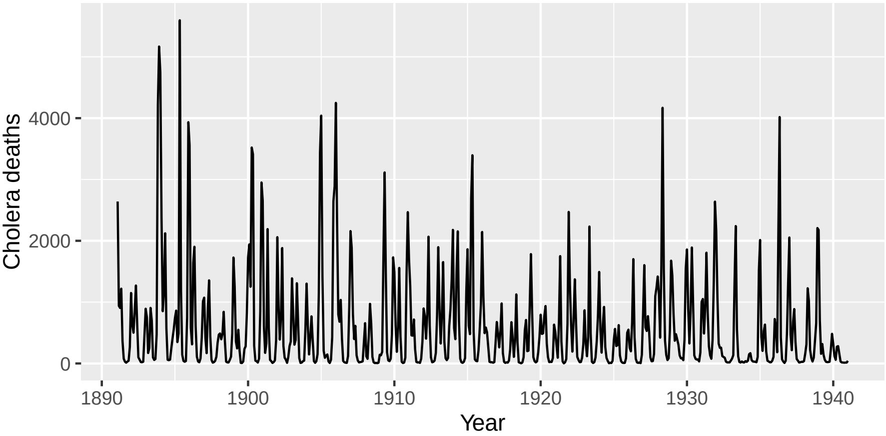
The nonlinearity arises through the force of infection, $\lambda(t)$, specified as

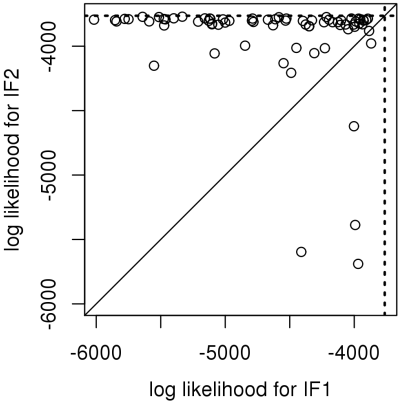
$$\lambda(t) = \bar{\beta} \exp\Big\{\beta_{\text{trend}}(t - t_0) + \textstyle\sum_{j=1}^{N_s} \beta_j s_j(t)\Big\} I/P + \bar{\omega} \exp\Big\{\textstyle\sum_{j=1}^{N_s} \omega_j s_j(t)\Big\},$$

where $\{s_j(t), j = 1, \ldots, N_s\}$ is a periodic cubic B-spline basis. The data are monthly counts of cholera mortality, modeled as

$$Y_n \sim \mathrm{Normal}(M_n, \tau^2 M_n^2) \text{ for } M_n = \int_{t_{n-1}}^{t_n} m\, I(s)\, ds.$$

Monthly cholera mortality in Dhaka, 1891-1940





Comparison of IF1 and IF2 on the cholera model.

Algorithmic tuning parameters for both IF1 and IF2 were set at the values chosen by King et al (2008) for IF1.

- Log likelihoods of the parameter vector output by IF1 and IF2, both started at a uniform draw from a large 23-dimensional hyper-rectangle.
- Dotted lines show the maximum log likelihood.

IF2 as an iterated Bayes map

- Each iteration of IF2 is a Monte Carlo approximation to a map

$$T_\sigma f(\theta_N) = \frac{\int \breve{\ell}(\theta_{0:N}) h(\theta_{0:N} | \varphi\, ; \sigma) f(\varphi)\, d\varphi\, d\theta_{0:N-1}}{\int \breve{\ell}(\theta_{0:N}) h(\theta_{0:N} | \varphi\, ; \sigma) f(\varphi)\, d\varphi\, d\theta_{0:N}}, \tag{1}$$

  where $\breve{\ell}(\theta_{0:N})$ is the likelihood of the data under the extended model with time-varying parameter $\theta_{0:N}$.
- f and $T_\sigma f$ in (1) approximate the initial and final density of the IF2 parameter swarm.
- When the standard deviation of the parameter perturbations is held fixed at $\sigma_m = \sigma > 0$, IF2 is a Monte Carlo approximation to $T_\sigma^M f(\theta)$.
- Iterated Bayes maps are not usually contractions.
- We study the homogeneous case, $\sigma_m = \sigma$.
- Studying the limit $\sigma \to 0$ may be as appropriate as an asymptotic analysis to study the practical properties of a procedure such as IF2, with σ_m decreasing down to some positive level $\sigma > 0$ but never completing the asymptotic limit $\sigma_m \to 0$.

IF2 as a generalization of data cloning

- In the case $\sigma = 0$, the iterated Bayes map corresponds to the data cloning approach of Lele et al (2007).
- For $\sigma = 0$, Lele et al (2007) found central limit theorems. For $\sigma \neq 0$, the limit as $M \to \infty$ is not usually Gaussian.
- Taking $\sigma \neq 0$ adds numerical stability, which is necessary for convergence of SMC approximations.

Theorem 1. Assuming adequate regularity conditions, there is a unique probability density f_σ with

$$\lim_{M\to\infty} T_\sigma^M f = f_\sigma,$$

with the limit taken in the L^1 norm. The SMC approximation to $T_\sigma^M f$ converges to $T_\sigma^M f$ as $J \to \infty$, uniformly in M.

- Theorem 1 follows from existing results on filter stability.
- Convergence and stability of the ideal filter (a small error at time t has diminishing effects at later times) is closely related to convergence of SMC.

Theorem 2. Under regularity conditions, $\lim_{\sigma \to 0} f_\sigma$ approaches a point mass at the maximum likelihood estimate (MLE).

Outline of proof.

- Trajectories in parameter space which stray away from the MLE are down-weighted by the Bayes map relative to trajectories staying close to the MLE.
- As σ decreases, excursions any fixed distance away from the MLE require an increasing number of iterations and therefore receive an increasing penalty from the iterated Bayes map.
- Bounding this penalty proves the theorem.

Conclusions

- IF1 enabled previously unfeasible likelihood-based inference for nonlinear, non-Gaussian POMP models.
- We have not yet found a situation where IF2 performs worse than IF1. In complex nonlinear models, we have found IF2 always substantially better.
- In addition, IF2 is simpler. Some extensions are easier: IF2 can readily handle parameters for which the information in the data is concentrated in a sub-interval.
- **IF2, together with advances in software and hardware, makes inference for nonlinear POMP models readily accessible to Masters level statisticians:**
    1. A short coure on simulation-based inference for epidemiological dynamics (`http://kingaa.github.io/sbied`).
    2. A course on time series analysis from a likelihood-based perspective (`https://github.com/ionides/531w16`)

References

Dobson, A. (2014). Mathematical models for emerging disease. *Science*, 346:1294–1295.

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.

Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., and King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences of the USA*, 112:719–724.

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

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43.

Kitagawa, G. (1998). A self-organising state-space model. *Journal of the American Statistical Association*, 93:1203–1215.

Lele, S. R., Dennis, B., and Lutscher, F. (2007). Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10:551–563.

Liu, J. and West, M. (2001). Combining parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–224. Springer, New York.