Solution uncertainty quantification for differential equations

Oksana A. Chkrebtii
Department of Statistics, The Ohio State University

Coauthors: David A. Campbell, Ben Calderhead, Mark Girolami

Parameter Estimation and Uncertainty Quantification for Dynamical Systems, University of Pittsburgh, March 5, 2017
Inference for dynamical systems

Dynamics of system states are often understood as functions of their rates of change with respect to spatio-temporal variables.

We often wish to infer parameters and system states that are defined implicitly, but observed directly.
Probability model for the data

We wish to estimate the unknowns, $\theta \in \Theta$, from the observations,

$$y(x_t) = A \{u(x_t, \theta)\} + \varepsilon(x_t), \quad x_t \in \mathcal{X}, \quad t = 1, \ldots, T,$$

of the deterministic state $u_t = u(x_t, \theta)$ transformed via an observation process $A$, and contaminated with stochastic noise $\varepsilon$.

Discrepancy between the model and the data is given by the likelihood,

$$f(y_{1:T} \mid \theta) \propto \rho \{y_{1:T} - A(u_{1:T})\}$$
The Bayesian formalism

Bayesian inference is concerned with modeling degree of belief about an unknown quantity via probability statements.

For example, we may not know $\theta \in \Theta$ but we may have some prior knowledge about, e.g., its range, most probable values, even before any data is collected,

$$\theta \sim \pi(\theta)$$

We seek to update our prior knowledge by conditioning on any new information, $y_{1:T} \in \mathcal{Y}$, e.g., data, model evaluations, via Bayes’ Rule,

$$p(\theta \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta) \pi(\theta)}{\int p(y_{1:T} \mid \theta) \pi(\theta) d\theta} \propto p(y_{1:T} \mid \theta) \pi(\theta)$$
Typical inverse problem formulation

The likelihood function involves the explicit solution, \( u_t = u(x_t, \theta) \), of the differential equation, typically unknown in closed form.

The classical approach replaces \( u \) with an approximate numerical solution, \( u^N \), to construct a surrogate hierarchical model, where \( N \) denotes the size the discretization grid or finite element mesh,

\[
\begin{align*}
[y_{1:T} \mid u_{1:T}, \theta] & \propto \rho[y_{1:T} - A(u_{1:T})] \\
[u_{1:T} \mid \theta] & \sim \delta(u_{1:T} - u^N_{1:T}(\theta)) \\
[\theta] & \sim \pi(\theta).
\end{align*}
\]

Estimators of \( \theta \) are typically posterior functionals, e.g., posterior mean, posterior mode, that can be approximated from a posterior sample.
The need for numerical uncertainty quantification
Example - galaxy simulation

AGORA
A High-resolution Galaxy Simulations Comparison Initiative: www.AGORAsimulations.org

High-res Galaxy Simulations

AGORA Comparison Infrastructure

AGORA Goal & Team

- GOAL: A collaborative, multi-platform study to raise the realism and predictive power of galaxy formation simulations

- TEAM: 115 participants from 60 institutions worldwide, Oct. 2014

- DATA SHARING: Simulations outputs and analysis softwares will be shared with the community

Contact: santacruzgalaxy@gmail.com

AGORA First Light: Flagship paper by Ji-hoon Kim et al. (ApJS 2014)

Project funded in part by:

THE OHIO STATE UNIVERSITY
Example - galaxy simulation
Example - galaxy simulation

- These are *not* realizations of a field (the model is deterministic!)
- The initial conditions and inputs are held fixed
- So which of these is our model \(u(x, \theta)\)?
A fully Bayesian inverse problem

We could aim for exact inference if we could model our uncertainty about the exact solution (which is unknown in closed form) given a discretization grid of size $N$.

\[
[y_{1:T} | u_{1:T}, \theta] \propto \rho[y_{1:T} - A(u_{1:T})]
\]

\[
[u_{1:T} | \theta] \sim \text{a probability measure representing uncertainty in the solution given discretization of size } N
\]

\[
[\theta] \sim \pi(\theta).
\]

How to define this middle layer?
A Probability Model for Discretization Uncertainty
Bayesian formalism for the unknown solution

Consider the initial value problem,

\[
\begin{cases}
Du = f(x, u), & x \in \mathcal{X}, \\
u = u_0 & x \in \partial \mathcal{X}
\end{cases}
\]

We may have some prior knowledge about smoothness, boundary conditions, etc., described by a prior probability with density,

\[u(x) \sim \pi(u(x)), \quad x \in \mathcal{X}\]

We seek to update our prior knowledge by conditioning model interrogations, \( f_n := f(x, u) \) via Bayes’ Rule,

\[
p(u(x) \mid f_{1:N}) = \frac{p(f_{1:N} \mid u(x)) \pi(u(x))}{\int p(f_{1:N} \mid u(x)) \pi(u(x)) \, du(x)} \propto p(f_{1:N} \mid u(x)) \pi(u(x))
\]
Prior uncertainty in the unknown solution

The exact solution function $u$ is deterministic, but unknown. We may describe our prior uncertainty via a probability model defined on the space of suitably smooth derivatives, e.g.,

$$u \sim \mathcal{GP}(m^0, C^0), \quad m^0 : \mathcal{X} \to \mathbb{R}, \quad C^0 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

satisfying the constraint $m^0 = u^0, x \in \partial \mathcal{X}$.

This yields a joint prior on the fixed but unknown state and its derivative(s)

$$
\begin{bmatrix}
  u \\
  Du
\end{bmatrix} \sim \mathcal{GP}
\left(
\begin{bmatrix}
  m^0 \\
  Dm^0
\end{bmatrix},
\begin{bmatrix}
  C^0 & C^0 D^* \\
  D C^0 & D C^0 D^*
\end{bmatrix}
\right)
$$
Interrogating the model recursively

1. Draw a sample from the marginal predictive distribution on the state at the next discretization grid point \( s_{n+1} \in \mathcal{X}, 1 \leq n < N \)

\[
u(s_{n+1}) \sim p(u(s_{n+1}) \mid f_{1:n})
\]

2. Evaluate the RHS at \( u(s_{n+1}) \) to obtain a model interrogation,

\[
f_{n+1} = f(s_{n+1}, u(s_{n+1}))
\]

3. Model interrogations as “noisy” measurements of \( Du \):

\[
f_{n+1} \mid Du, f_{1:n} \sim \mathcal{N}(Du(s_{n+1}), \Lambda(s_n))
\]
Sequential Bayesian updating

Updating our knowledge about the true but unknown solution given the new interrogation trajectory $f_{n+1}$

$$
\begin{bmatrix}
u \\
Du
\end{bmatrix} | f_{n+1} \sim \mathcal{GP}
\left(\begin{bmatrix} m^{n+1} \\
Dm^{n+1}\end{bmatrix}, \begin{bmatrix} C^{n+1} & C^{n+1}D^* \\
DC^{n+1} & DC^{n+1}D^* \end{bmatrix}\right)
$$

where,

$$
\begin{align*}
m^{n+1} &= m^n + K^n (f_{n+1} - m^n(s_{n+1})) \\
C^{n+1} &= C^n - K^n DC^{n*} \\
K^n &= C^n D^* (DC^n + \Lambda(s_n))^{-1}
\end{align*}
$$

This becomes the prior for the next update.
“Probabilistic Solution” model of discretization uncertainty

Due to the Markov property, we cannot condition the solution on multiple trajectories $f_1, \ldots, f_N$ simultaneously. Instead, the posterior is a continuous mixture of Gaussian processes as follows,

$$[u \mid \theta, \Psi, N] = \int \int [u, Du \mid f_{1:N}, \theta, \Psi, N] \, d(Du) \, df_{1:N}$$

Draws from this posterior can be obtained via Monte Carlo.
Example - a simple ODE initial value problem

\[
\begin{aligned}
\frac{d^2}{dx^2} u(x) &= \sin(2x) - u(x), \quad x \in [0, 10], \\
\frac{d}{dx} u(0) &= 0, \quad u(0) = -1.
\end{aligned}
\]

This problem has a closed form solution (red), assumed unknown a-priori.

*Five draws from the prior process for the state (left) and first derivative (right)*
Example - a simple ODE initial value problem

\[
\begin{align*}
\frac{d^2}{dx^2} u(x) &= \sin(2x) - u(x), \quad x \in [0, 10], \\
\frac{d}{dx} u(0) &= 0, \quad u(0) = -1.
\end{align*}
\]

This problem has a \textbf{closed form solution} (red), assumed unknown a-priori.

*Five draws from the prior process for the state (left) and first derivative (right)*
Example - a simple ODE initial value problem

\[
\begin{align*}
\frac{d^2}{dx^2} u(x) &= \sin(2x) - u(x), \quad x \in [0, 10], \\
\frac{d}{dx} u(0) &= 0, \quad u(0) = -1.
\end{align*}
\]

This problem has a closed form solution (red), assumed unknown a-priori.

Five draws from the prior process for the state (left) and first derivative (right)
Example - a simple ODE initial value problem

\[
\begin{align*}
\frac{d^2}{dx^2} u(x) &= \sin(2x) - u(x), \quad x \in [0, 10], \\
\frac{d}{dx} u(0) &= 0, \quad u(0) = -1.
\end{align*}
\]

This problem has a closed form solution (red), assumed unknown a-priori.

*Five draws from the prior process for the state (left) and first derivative (right)*
Example - a simple ODE initial value problem

\[
\begin{cases}
\frac{d^2}{dx^2} u(x) = \sin(2x) - u(x), & x \in [0, 10], \\
\frac{d}{dx} u(0) = 0, & u(0) = -1.
\end{cases}
\]

This problem has a closed form solution (red), assumed unknown a-priori.

Five draws from the prior process for the state (left) and first derivative (right)
Example - a simple second order ODE model

A probability model for uncertainty in the unknown solution provides a trade-off between accuracy and computational expenditure (grid refinement)

Exact solution (green); Given $N = 50, 100, 200$ (left to right) 100 draws (grey) from probability model of solution uncertainty, numerical solution (red)
Example - Lorenz63 forward model

A probability statement over probable trajectories given fixed model parameters and initial conditions for the Lorenz63 model:

1000 draws for the probabilistic forward model for the Lorenz63 system given fixed initial states and model parameters in the chaotic regime.
Example - Lorenz63 forward model

1000 draws from forward model for Lorenz63 system at four fixed time points.
Fully Bayesian Inverse Problem
Fully Bayesian Inverse Problem

We can now perform exact inference with solution uncertainty based on the following hierarchical model,

\[
[y_{1:T} \mid u_{1:T}, \theta] \propto \rho [y_{1:T} - A(u_{1:T})]
\]

\[
[u_{1:T} \mid \theta] \sim \int [u(s_{1:N}, \theta) \mid f_{1:N}, \theta] \, df_{1:N}
\]

\[
[\theta] \sim \pi(\theta).
\]

Next, let us consider an example from systems biology
Inference for a model of protein dynamics

JAK-STAT chemical signaling pathway model describes concentration of 4 STAT factors by a delay differential equation system on $t \in [0, 60]$,

\[
\begin{align*}
\frac{d}{dt} u^{(1)}(t, \theta) &= -k_1 u^{(1)}(t, \theta) EpoR_A(t) + 2k_4 u^{(4)}(t - \tau, \theta) \\
\frac{d}{dt} u^{(2)}(t, \theta) &= k_1 u^{(1)}(t, \theta) EpoR_A(t) - k_2 \left(u^{(2)}(t, \theta)\right)^2 \\
\frac{d}{dt} u^{(3)}(t, \theta) &= -k_3 u^{(3)}(t, \theta) + 0.5k_2 \left(u^{(2)}(t, \theta)\right)^2 \\
\frac{d}{dt} u^{(4)}(t, \theta) &= k_3 u^{(3)}(t, \theta) - k_4 u^{(4)}(t - \tau, \theta)
\end{align*}
\]

\[
u^{(i)}(t, \theta) = \phi^{(i)}(t), \quad t \in [-\tau, 0], \; i = 1, \ldots, 4
\]
Inference for a model of protein dynamics

States are observed indirectly through a nonlinear transformation:

\[ A^{(1)} = k_5 (u^{(1)}(t; \theta) + 2u^{(3)}(t; \theta)) \]
\[ A^{(2)} = k_6 (u^{(1)}(t; \theta) + u^{(2)}(t; \theta) + 2u^{(3)}(t; \theta)) \]
\[ A^{(3)} = u^{(1)}(t; \theta) \]
\[ A^{(4)} = \frac{u^{(3)}(t; \theta)}{u^{(2)}(t; \theta) + u^{(3)}(t; \theta)} \]

Observations are noisy measurements on the transformed states and forcing function at points \( t = \{ t_{ij} \}_{i=1,...,4;j=1,...,n_i} \)

\[ y(t) = A_{k_4,k_5} u(t; k_1, \ldots, k_6, \tau, \phi, EpoR_A) + \varepsilon(t) \]
Inference for a model of protein dynamics

Sample from the marginal posterior distribution over the states transformed via the observation process (gray); experimental measurements (red)
Inference for a model of protein dynamics

Marginal fully probabilistic posterior distribution in the model parameters based on a sample of size 100,000 generated by a parallel tempering algorithm utilizing seven chains, with the first 10,000 samples removed. Prior densities are shown in red.
Ongoing and future work

- Interrogating an emulated forward model greatly reduces computational cost for the full inverse problem

- Incorporating boundary constraints in the prior model on the solution requires complex spatial probability models, potentially sacrificing closed form updates

- Adaptive mesh selection for probabilistic numerical solvers

- Probabilistic finite elements solvers
Thank you!

References:

1. Pratola, M.T., Chkrebtii, O.A. *Bayesian Calibration of Multistate Stochastic Simulators*. In revision.


Models and Examples:


Contact: oksana@stat.osu.edu