Sequential Monte Carlo for estimating parameters of differential equations

Liangliang Wang

Dept. of Statistics and Actuarial Science
Simon Fraser University (SFU)
Burnaby, BC, Canada

Oct 16, 2020
<table>
<thead>
<tr>
<th></th>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Methods for estimating DE parameters</td>
</tr>
<tr>
<td>3</td>
<td>Bayesian inference via sequential Monte Carlo</td>
</tr>
<tr>
<td>4</td>
<td>Simulation</td>
</tr>
<tr>
<td>5</td>
<td>Application</td>
</tr>
<tr>
<td>6</td>
<td>Summary</td>
</tr>
</tbody>
</table>
Population of Blowflies


- The blowflies were cultured in a room maintained at 25°C.

- Nicholson measured the population of blowflies every day for approximately one year.
Number of Blowflies VS Days

- Counts of Blowflies
- Time

Liangliang Wang (SFU)
SMC for DE parameters
Oct 16, 2020
Population of Blowflies

\[ \frac{dx(t)}{dt} = r x(t) \left( 1 - \frac{x(t - \tau)}{1000 \cdot P} \right) \]

- \( x(t) \): the blowfly population
- \( r \): the rate of increase of the blowfly population
- \( P \): a resource limitation parameter set by the supply of food
- \( \tau \): the time delay, roughly equal to the time for a larva to grow up to an adult
Delay Differential Equations (DDEs)

A typical form for DDEs:

\[
\frac{dx_i(t)}{dt} = g_i(x(t), x(t - \tau)|\theta),
\]

\[
x_i(0) = x_{i0},
\]

- $x(t)$: the dynamic process on $[t_1, t_J]$.
- $\tau$, a constant delay parameter.
- $\theta$, a vector of parameters describing the system.
- The parametric form of $g_i(\cdot)$ is often known.
- Ordinary Differential Equations (ODE) is a special case of DDE when $\tau = 0$. 

Likelihood function

- Let $y_{ij}$ be the $j$-th observation of the $i$-th DE $x_i(\cdot)$ at time $t_{ij}$.

- Likelihood function:

  $$L(y_{ij}|x_i(t_{ij}), \theta_y), j = 1, \ldots, J,$$

  where $\theta_y$ is the vector of the parameters in the observation model.

- For example,

  $$y_{ij} \sim N(x_i(t_{ij}), \sigma_i^2), j = 1, \ldots, J,$$

  where $\theta_y = (\sigma_1^2, \ldots, \sigma_I^2)^T$. 
Challenges

Question: Estimate $\theta$, $\tau$, and $x_{i0}$ from noisy and/or partially observed data?

Challenges

- The DDEs usually have no analytic solutions and can only be solved numerically.

- The DDE solutions will not only depend on the parameters $\theta$ and $\tau$, but will also rely on the history of the dynamic process $\mathcal{H}_\tau = \{x(t), t \in [t_1 - \tau, t_1]\}$ which is an infinite-dimensional set.

- DDE solutions are extremely sensitive to delay parameters.
Methods

Outline

1. Introduction
2. Methods for estimating DE parameters
3. Bayesian inference via sequential Monte Carlo
4. Simulation
5. Application
6. Summary
Maximum Likelihood Method

The DE parameters $\theta$ can be estimated by maximizing the log likelihood function

$$H^*(\theta) = \sum_{i=1}^{n} \log L(y_i|X(t_i|\theta)).$$

Difficulties with MLE

- The DE solution $X(t|\theta)$ is an implicit function of $\theta$.
- High computation load associated with the numeric solver.
- Many local optima in the log likelihood function.
- We have to specify the unknown history of the dynamic process $\mathcal{H}_T$ to solve the DDE numerically.
A two-step procedure

Ellner et al. (1997)

**Step 1: Nonparametric Smoothing**

- \( y_i = X(t_i) + \epsilon_i \)
- Estimate \( X(t) \) and \( X'(t) \) from noisy data using nonparametric smoothing methods

**Step 2: Nonlinear Regression**

\[
\min_{\theta, \tau} \sum_{i=1}^{n} \| \hat{X}'(t_i) - g(\hat{X}(t), \hat{X}(t - \tau)|\theta) \|^2
\]
Pros and cons of the two-step procedure

Pros
- No need to solve DDE
- Fast computation

Cons
- $X'(t)$ may not be well estimated, especially from sparse data.
- Thus DE parameter estimation is not accurate.
Semiparametric Method

Wang and Cao (2012)

Advantages

- No need to solve DDE numerically
- No need to specify DDE history
- Estimation is more accurate than the two-step method.
Semiparametric Method

Estimate a nonparametric function to approximate DDE solutions by a linear combination of basis functions:

\[ x(t) = \sum_{k=1}^{K} \phi_k(t)c_k = \phi^T(t)c, \]

- \( \phi(t) = (\phi_1(t), \ldots, \phi_K(t))^T \) is the vector of basis functions at time \( t \), for example, B-spline (Fixed and Known).
- \( c = (c_1, \ldots, c_K)^T \) are the basis function coefficients.
Cubic B-spline Basis
Estimating Spline Coefficients

Fitting to data
- Observations: \( y_i \)
- Nonparametric function: \( x(t) = \phi^T(t) \mathbf{c} \)
- Fitting to data: \( C_1 = \sum_{i=1}^{n} \log L(y_i | x(t_i)) \)

Fidelity to DDE
\[
\frac{dx(t)}{dt} = g(x(t), x(t - \tau) | \theta)
\]

- Difference between two sides of DDE: \( \frac{dx(t)}{dt} - g(x(t), x(t - \tau) | \theta) \)
- Fidelity to DDE: \( C_2 = \int_{t_1 + \tau}^{t_n} \left( \frac{dx(t)}{dt} - g(x(t), x(t - \tau) | \theta) \right)^2 dt \)

Criterion: \( J(\mathbf{c} | \theta) = C_1 + \lambda C_2 \)
Estimating Spline Coefficients

Fitting to data

- Observations: $y_i$
- Nonparametric function: $x(t) = \phi^T(t)c$
- Fitting to data: $C_1 = \sum_{i=1}^{n} \log L(y_i|x(t_i))$

Fidelity to DDE $\frac{dx(t)}{dt} = g(x(t), x(t-\tau)|\theta)$

- Difference between two sides of DDE: $\frac{dx(t)}{dt} - g(x(t), x(t-\tau)|\theta)$
- Fidelity to DDE: $C_2 = \int_{t_1+\tau}^{t_n} \left\{ \frac{dx(t)}{dt} - g(x(t), x(t-\tau)|\theta) \right\}^2 dt$

Criterion: $J(c|\theta) = C_1 + \lambda C_2$
Estimating Spline Coefficients

**Fitting to data**
- Observations: $y_i$
- Nonparametric function: $x(t) = \phi^T(t)c$
- Fitting to data: $C_1 = \sum_{i=1}^{n} \log L(y_i|x(t_i))$

**Fidelity to DDE** $\frac{dx(t)}{dt} = g(x(t), x(t - \tau)|\theta)$

- Difference between two sides of DDE: $\frac{dx(t)}{dt} - g(x(t), x(t - \tau)|\theta)$
- Fidelity to DDE: $C_2 = \int_{t_1+\tau}^{t_n} \left( \frac{dx(t)}{dt} - g(x(t), x(t - \tau)|\theta) \right)^2 dt$

**Criterion:** $J(c|\theta) = C_1 + \lambda C_2$
Estimating DDE parameter $\theta$

- **Criterion:**

$$H(\theta) = \sum_{i=1}^{n} \log L(y_i|\hat{x}(t_i|\theta)),$$

where $\hat{x}(t|\theta) = \phi^T(t)\hat{c}(\theta)$ and we obtain the estimate $\hat{c}$ by maximizing the penalized log likelihood function.
Two nested levels of optimization

- **Inner level:** $J(c|\theta)$
  - $\hat{c}$ is a function of $\theta$: $\hat{c}(\theta)$.

- **Outer level:** $H(\hat{c}(\theta), \theta)$

$c$: basis coefficients; $\theta$: DDE parameter;
Diagram for our semiparametric method
The $K$-fold cross validation:

$$CV = \frac{1}{K} \sum_{j=1}^{K} \sum_{i \in A(-j)} \log L(y_i | \hat{x}(t_i | \theta^{(-j)})).$$
Outline

1. Introduction
2. Methods for estimating DE parameters
3. Bayesian inference via sequential Monte Carlo
4. Simulation
5. Application
6. Summary
Posterior distribution

\[
\pi(\theta_{DE}, \theta_y, c, \lambda) \propto \gamma(\theta_{DE}, \theta_y, c, \lambda)
\]

\[
= L(y|\theta_y, c)\pi_0(c|\theta_{DE}, \lambda)\pi_0(\theta_{DE}, \theta_y, \lambda),
\]

where \(\gamma\) is the unnormalized posterior distribution, and \(\theta_{DE}\) includes \(\theta\) and \(\tau\).
Prior distribution for $c$

Given $\lambda, \theta, \tau$, the prior distribution for $c$ is

$$
\pi_0(c|\theta, \tau, \lambda) \propto \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^{l} \int_{t_i+\tau}^{t_{i+1}} \left[ \frac{dx_i(s)}{ds} - g_i(x(s), x(s-\tau)|\theta) \right]^2 ds \right\},
$$

$$
= \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^{l} \int_{t_i+\tau}^{t_{i+1}} \left[ \frac{d\Phi(s)'c_i}{ds} - g_i(\Phi(s)'c, \Phi(s-\tau)'c|\theta) \right]^2 ds \right\}.
$$

This prior distribution measures how well the estimated DE variables $\hat{x}(t)$ satisfy the DE system.
Prior distributions $\pi_0(\theta_{DE}, \theta_y, \lambda)$

For example,

\[ \theta \sim \text{MVN}(0_D, \sigma^2_\theta I_D), \]
\[ \tau \sim \text{U}(t_1, t_J), \]
\[ \sigma^2_i \sim \text{IG}(g_0, h_0), \quad i = 1, \ldots, I, \]
\[ \lambda \sim \text{Gamma}(a_\lambda, b_\lambda). \]
Likelihood function

For example,

\[ L(y|c, \sigma) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \sigma_i^2 \]^{-1/2} \exp \left\{ - \sum_{i=1}^{I} \left( \sum_{j=1}^{J} \frac{(y_{ij} - \Phi(t_{ij})'c_i)^2}{2\sigma_i^2} \right) \right\} \]
The $r$-th intermediate distribution:

\[
\pi_r(\theta_\text{DE}, \theta_y, c, \lambda) \propto \gamma_r(\theta_\text{DE}, \theta_y, c, \lambda) \\
= [L(y|\theta_y, c)\pi_0(c|\theta_\text{DE}, \lambda)\pi_0(\theta_\text{DE}, \theta_y, \lambda)]^{\alpha_r} \rho(\theta_\text{DE}, \theta_y, c, \lambda)^{1-\alpha_r}, \\
= [L(y|\beta)\pi_0(\beta)]^{\alpha_r} \rho(\beta)^{1-\alpha_r}
\]

- $\rho(\theta_\text{DE}, \theta_y, c, \lambda)$ is a reference distribution.
- $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_{R-1} < \alpha_R = 1$ is a sequence of annealing parameters.
- $\pi_0(\theta_\text{DE}, \theta_y, c, \lambda) = \rho(\theta_\text{DE}, \theta_y, c, \lambda)$.
- $\pi_R(\theta_\text{DE}, \theta_y, c, \lambda) = \pi(\theta_\text{DE}, \theta_y, c, \lambda)$.
- Denote $\beta = (\theta_\text{DE}, \theta_y, c, \lambda)^T$. 
Annealed SMC

(Del Moral et al. 2006, 2007)

sample $\tilde{\beta}_{0,k} \sim \rho(\cdot)$

set its unnormalized weight $w_{0,k} = 1$.

for $r \in 1, \ldots, R$ do

sample $\beta_{r,k} \sim T_r(\tilde{\beta}_{r-1,k}, \cdot)$

compute

$$w_{r,k} = w_{k,r-1} \cdot \frac{\gamma_r(\beta_{r,k})}{\gamma_{r-1}(\tilde{\beta}_{r-1,k})} \cdot \frac{L_{r-1}(\beta_{r,k}, \tilde{\beta}_{r-1,k})}{T_r(\tilde{\beta}_{r-1,k}, \beta_{r,k})}$$

normalize weights $W_{r,k} = w_{r,k} / \sum_{k=1}^{K} w_{r,k}$.

if $\text{ESS} < \text{threshold}$, resample $\{\beta_{r,k}, W_{r,k}\}$ to obtain new particles denoted $\{\tilde{\beta}_{r,k}\}$, and set $w_{r,k} = 1$.

end for
Markov kernels $T_r(\tilde{\beta}_{r-1}, \cdot)$

- We propagate new particles $\{\beta_{r,k}\}$ via $\pi_r$-invariant MCMC moves, $\beta_{r,k} \sim T_r(\tilde{\beta}_{r-1,k}, \cdot)$.

- Sample from the full conditional distributions:
  - $\pi_r(\theta_{DE}|\cdot)$
  - $\pi_r(\theta_y|\cdot)$
  - $\pi_r(c_i|\cdot)$
  - $\pi_r(\lambda|\cdot)$

- If the full conditional distribution does not admit a closed form, we implement one step of the Metropolis-Hastings algorithm.
Bayesian inference via SMC

Backward Markov kernel

A convenient backward Markov kernel that allows an easy evaluation of the importance weight is

$$L_{r-1}(\beta_r, \tilde{\beta}_{r-1}) = \frac{\pi_r(\tilde{\beta}_{r-1}) T_r(\tilde{\beta}_{r-1}, \beta_r)}{\pi_r(\beta_r)}.$$  

With this backward kernel, the incremental importance weight becomes

$$w_r = w_{r-1} \cdot \frac{\gamma_r(\beta_r)}{\gamma_{r-1}(\tilde{\beta}_{r-1})} \cdot \frac{L_{r-1}(\beta_r, \tilde{\beta}_{r-1})}{T_r(\tilde{\beta}_{r-1}, \beta_r)}$$

$$= w_{r-1} \cdot \frac{\gamma_r(\tilde{\beta}_{r-1})}{\gamma_{r-1}(\tilde{\beta}_{r-1})}$$

$$= w_{r-1} \cdot \left( \frac{L(y|\tilde{\beta}_{r-1})\pi_0(\tilde{\beta}_{r-1})}{\rho(\tilde{\beta}_{r-1})} \right)^{\alpha_r - \alpha_{r-1}}$$
Choose $\alpha_r - \alpha_{r-1}$ adaptively

Conditional ESS (CESS) (Zhou, Johansen, and Aston (2016))

\[
\text{CESS}_r = \frac{K(\sum_{k=1}^{K} W_{r-1,k} w_{r,k})^2}{\sum_{k=1}^{K} W_{r-1,k}(w_{r,k})^2},
\]

which is a function of $\alpha_r - \alpha_{r-1}$.

The next annealing parameter $\alpha_r$ can be determined by solving $\text{CESS}_r = \phi_r$, a user specified threshold. (Wang, Wang, and Bouchard-Côté (2020))
Simulation

Outline

1. Introduction
2. Methods for estimating DE parameters
3. Bayesian inference via sequential Monte Carlo
4. Simulation
5. Application
6. Summary
Simulation setting

- ODE

\[
\frac{dx_1(t)}{dt} = \frac{72}{36 + x_2(t)} - |\theta_1|,
\]
\[
\frac{dx_2(t)}{dt} = \theta_2 x_1(t) - 1.
\]

- Observation model: \( y_i \sim \text{Normal}(X(t_i|\theta), \sigma^2) \).

- \(|\theta_1| = 2, \theta_2 = 1\).

- \(x_1(0) = 7\), and \(x_2(0) = -10\).

- The standard deviation of the noise is set to \(\sigma = 0.5\).
ASMC results
## Comparison SMC-spline to alternative methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Target distribution</th>
<th>ODE solver</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMC-spline</td>
<td>$\pi(\theta_{DE}, \theta_y, c, \lambda)$</td>
<td></td>
<td>ASMC</td>
</tr>
<tr>
<td>SMC-deSolve</td>
<td>$\pi(\theta_{DE}, \theta_y, x(0))$</td>
<td>“lsoda”</td>
<td>ASMC</td>
</tr>
<tr>
<td>MCMC-spline</td>
<td>$\pi(\theta_{DE}, \theta_y, c, \lambda)$</td>
<td></td>
<td>MCMC</td>
</tr>
<tr>
<td>MCMC-deSolve</td>
<td>$\pi(\theta_{DE}, \theta_y, x(0))$</td>
<td>“lsoda”</td>
<td>MCMC</td>
</tr>
</tbody>
</table>
Comparison SMC-spline to alternative methods

(a) SMC1

(b) SMC2

(c) MCMC1

(d) MCMC2
## Comparison SMC-spline to alternative methods

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>SMC-spline</th>
<th>SMC-deSolve</th>
<th>MCMC-spline</th>
<th>MCMC-deSolve</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\theta_1</td>
<td>$</td>
<td>2</td>
<td>1.93 (1.68, 2.19)</td>
<td>1.84 (1.81, 1.88)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>1</td>
<td>0.99 (0.90, 1.09)</td>
<td>1.12 (1.08, 1.17)</td>
<td>0.97 (0.87, 1.08)</td>
<td>0.85 (0.85, 0.85)</td>
</tr>
<tr>
<td>$x_1(0)$</td>
<td>7</td>
<td>6.43 (2.56, 10.11)</td>
<td>3.94 (3.71, 4.22)</td>
<td>6.48 (4.01, 8.92)</td>
<td>4.55 (4.55, 4.55)</td>
</tr>
<tr>
<td>$x_2(0)$</td>
<td>-10</td>
<td>-10.66 (-17.86, -2.26)</td>
<td>-4.47 (-5.63, -3.23)</td>
<td>-10.28 (-14.18, -6.71)</td>
<td>0.68 (0.65, 0.70)</td>
</tr>
</tbody>
</table>
1. Introduction
2. Methods for estimating DE parameters
3. Bayesian inference via sequential Monte Carlo
4. Simulation
5. Application
6. Summary
DDE for Blowflies

\[ \frac{dX(t)}{dt} = rX(t) \left( 1 - \frac{X(t - \tau)}{1000 \cdot P} \right) \]

- \( X(t) \) is the blowfly population
- \( r \) is the rate of increase of the blowfly population
- \( P \) is a resource limitation parameter set by the supply of food
- \( \tau \) is the time delay, roughly equal to the time for a larva to grow up to an adult
Estimated posterior mean trajectory and 95% credible bands for the delay differential equation modelling the population dynamics of blowflies. Here $X(t) = e^{W(t)}$. 
Posterior samples of $r$, $\tau$, $P$ estimated via SMC. Correlation:
$$\text{corr}(r, \tau) = 0.139, \text{corr}(r, P) = 0.598, \text{corr}(P, \tau) = 0.008.$$ 

Estimate for $\tau$

- Annealed SMC:
  - Posterior mean and its 95% CI: $8.368$ ($5.656, 9.916$)

  - $\hat{\tau} = 8.78$, standard error $0.039$
Posterior samples of $r$, $\tau$, $P$ estimated via SMC. Correlation:
$\text{corr}(r, \tau) = 0.139$, $\text{corr}(r, P) = 0.598$, $\text{corr}(P, \tau) = 0.008$.

Estimate for $\tau$
- Annealed SMC:
  - Posterior mean and its 95% CI: 8.368 (5.656, 9.916)
  - $\hat{\tau} = 8.78$, standard error 0.039
Outline

1. Introduction
2. Methods for estimating DE parameters
3. Bayesian inference via sequential Monte Carlo
4. Simulation
5. Application
6. Summary
Many statistical challenges for estimating ODE/DDE parameters.

Our Bayesian semiparametric method via SMC has several advantages:

- No need to solve DEs numerically
- No need to specify DDE history
- Provides uncertainty estimation for the parameters
- Shows advantages over MCMC methods
Acknowledgements

Shijia Wang (Nankai University)

The Natural Sciences and Engineering Research Council of Canada (NSERC).

Thank you!


