Parametric Bayesian Approach to Time Delay Estimation

Hyungsuk Tak

Stat 310

1 April 2014
Introduction

Time delay challenge
  - Data description
  - Complexities and Challenges

Popular estimation methods

Bayesian Approach
  - Idea and model specification
  - Prior distribution: Ornstein-Uhlenbeck process
  - Hyper-prior distribution
  - Full posterior distribution
  - Conditional posterior distributions
  - Two possible samplers for one light curve

Two real data examples
  - Data from Burud et al. (2002)
  - Data from Kochanek et al. (2006)

Discussion
Light travels via different paths due to the gravitational fields of intervening matter

- Several paths cause **multiple images of the same source.**
- **Different path lengths** cause different arrival times.
Introduction

Two light curves for a simulated double lensed quasar from the Time Delay Challenge (TDC) design paper (Dobler et al. 2013)

- Blue light curve lags behind the orange light curve as a result of the gravitational time delay
Time Delay Challenge

Accurate time delay estimate is important in

- measuring **cosmological parameters**, e.g., Hubble constant, $H_0$
- probing the **dark matter (sub-)structure** within the lens galaxy

Evil team gave a simulated data set (TDC0) to Good team.

- TDC0, called a ladder, consists of **7 rungs** (increasing difficulty).

- Each rung (subscript $j$) has **8 data sets** (subscript $i$).
- Each data set contains a pair of light curves with measurement errors.
- Good team’s job is to estimate the time delays in each dataset, $\hat{\Delta}_{ij}$, where $i = 1, 2, \ldots, 8$ and $j = 1, 2, \ldots, 7$. 
Data Description

5 variables in each dataset

- \textit{time}: observation (arrival) time in days
- \textit{lcA}: Intensity of the leading light curve A (red curve below) in nanomaggies
- \textit{se.lcA}: measurement error of the leading light curve A
- \textit{lcB}: Intensity of the following light curve B (blue curve below) in nanomaggies
- \textit{se.lcB}: measurement error of the following light curve B
**Complexities and Challenges**

- **Rung1 → Rung2**: Seasonal gaps
- **Rung2 → Rung3**: More variations in the following blue light curve
- **Rung1 → Rung4**: Sparse (irregular) observations (sampling time)
- **Rung4 → Rung5**: Seasonal gaps
- **Rung5 → Rung6**: Non-interger time (sampling time on real line)
- **Rung6 → Rung7**: More variations in the following blue light curve
Popular Estimation Methods

- **Smoothing and $\chi^2$-minimization (Fassnacht, 1999)**
  - **Smooth** both light curves
  - **Scale** (by $\sigma$) and **shift** (by $\Delta$) one smooth light curve
  - Calculate $\chi^2_{\sigma,\Delta}$ statistic
  - Find $\Delta$ minimizing $\chi^2_{\sigma,\Delta}$ on the two-dimensional grids of $\sigma$ and $\Delta$

- **Smoothing and Cross-correlation (Fassnacht, 1999)**
  - **Smooth** both light curves
  - **Shift** (by $\Delta$) one smooth light curve
  - Calculate $r_\Delta$, sample cross-correlation functions
  - Find $\Delta$ that maximizes $r_\Delta$ on the grid of $\Delta$
Popular Estimation Methods

- Dispersion method (Pelt et al. 1994)
  - Does not smooth the curves at all.
  - Introduce the composite curve merging two light curves, $X(t)$ and $Y(t + \Delta) + c$.
  - Calculate the dispersion ($D_{c,\Delta}^2$), defined as the weighted sum of squared differences of two adjoining points of the composite curve.
  - Find $\Delta$ minimizing $D_{c,\Delta}^2$ on the two-dimensional grids of $c$ and $\Delta$.

- Gaussian process (GP) (Tewes et al. 2013, Hojjati et al. 2013)
  - Fit the GPs on $X(t)$ and $Y(t)$ ($GP1$ and $GP2$ each), estimating the mean functions given certain covariance kernels.
  - Find $\Delta$ that minimizes the weighted average variation of difference curve, $GP1(t) - GP2(t + \Delta)$, on grid of $\Delta$. 
Idea and Model Specification

- ∃ only one underlying light curve: one light curve is just a shifted version of the other in $x$- and $y$-axes, i.e. $Y(t) = X(t - \Delta) + c$.

- **SNoTE:**

  - Two underlying light curves (blue is shifted version of red) with three observations each
    
    \[
    X(t) = \sin(t) \\
    Y(t) = X(t - 2) + 1
    \]

    - Blue curve lags behind Red one by 2 days, shifted by 1 unit in $y$-axis
    - 3 observations from each curve at $t_1$, $t_2$, and $t_3$
    - Time sequence on the Red light curve corresponding to six observations: $(t_1 - \Delta, t_1, t_2 - \Delta, t_3 - \Delta, t_2, t_3)$
Idea and Model Specification

Underlying process: $X(t_1 - \Delta) + c \rightarrow X(t_1) \rightarrow X(t_2 - \Delta) + c \rightarrow X(t_3 - \Delta) + c \rightarrow X(t_2) \rightarrow X(t_3)$

Observations: $y(t_1) \rightarrow x(t_1) \rightarrow y(t_2) \rightarrow y(t_3) \rightarrow x(t_2) \rightarrow x(t_3)$

- **Likelihood:**
  \[
  \begin{align*}
  x(t_j) &= X(t_j) + \epsilon_j, \quad \epsilon_j \sim N(0, \delta_j^2), \\
  y(t_j) &= X(t_j - \Delta) + c + \epsilon_j, \quad \epsilon_j \sim N(0, \eta_j^2), \quad j = 1, 2, \ldots, n
  \end{align*}
  \]

- **Prior**
  - $p(X(t), X(t - \Delta)|\theta, \Delta) = p(X(t')|\theta, \Delta)$, where $t' \equiv (t_1', t_2', \ldots, t_n') \equiv \text{sort}(t_1, t_2, \ldots, t_n, t_1 - \Delta, t_2 - \Delta, \ldots, t_n - \Delta)$
  - $p(\theta, \Delta, c)$

- **Hyper-prior**
Prior: Ornstein-Uhlenbeck Process

- Need to build a model for underlying (latent) light curve
  \[ p(X(t), X(t - \Delta)|\theta, \Delta) = p(X(t')|\theta, \Delta) \]
  - Stochastic process in continuous time
  - Easy way to sample light curve at irregularly-spaced times
- O-U process, also called CAR(1) or damped random walk process
- \[ dX(t) = -\frac{1}{\tau} (X(t) - \mu) dt + \sigma dB(t), \text{ where} \]
  - \( \tau \) is a relaxation time, \( \mu \) and \( \sigma \) are mean and scale parameters of the underlying process, and finally \( B(t) \) is a standard Brownian motion.
- Solution of stochastic differential equation with Marknovian property
  \[ X(t_j)|X(t_{j-1}), \mu, \sigma^2, \tau \sim N[\text{mean: } \mu + e^{-(t_j-t_{j-1})/\tau} (X(t_{j-1}) - \mu), \text{ variance: } \frac{\tau\sigma^2}{2} (1 - e^{-2(t_j-t_{j-1})/\tau})] \]
- \[ p(X(t')|\theta, \Delta) = p(X(t'_1)|\theta, \Delta) \prod_{j=2}^{2n} p(X(t'_j)|X(t'_{j-1}), \theta, \Delta) \]
Hyper-prior Distribution

- 5 hyper-parameters:
  - $\mu$ is a mean parameter of underlying process
  - $\sigma$ is a scale parameter of underlying process
  - $\tau$ is a relaxation time of the underlying process
  - $c$ is a shift in y-axis
  - $\Delta$ is a shift in x-axis (time delay)

- Naively informative: $p(\theta, c, \Delta) \equiv p(\mu, \sigma^2, \tau, c, \Delta) \propto \frac{1}{\sigma} \frac{e^{-\epsilon_1/\tau}}{\tau^{\epsilon_1+1}} \frac{e^{-\epsilon_2/\Delta}}{\Delta^{\epsilon_2+1}}$

- $\tau \sim InvGam(\epsilon_1, \epsilon_1)$ and $\Delta \sim InvGam(\epsilon_2, \epsilon_2)$

- In general, a diffuse hyper-prior distribution (possibly Normal) on $\Delta$, if we do not know which light curve is preceding
Full Posterior Distribution

- Full Posterior: \( p(X(t), X(t-\Delta), \theta, c, \Delta|x(t), y(t)) \)
  \[ \propto p(x(t)|X(t)) \cdot p(y(t)|X(t-\Delta) + c, c) \cdot p(X(t), X(t-\Delta)|\theta, \Delta) \cdot p(\theta, c, \Delta) \]

- Kelly et al. (2009) introduces a way to obtain a marginalized posterior distribution \( p(\theta, c, \Delta|x(t), y(t)) \) with the underlying process, \( X(t) \) and \( X(t-\Delta) \), integrated out.
Conditional Posterior Distributions

- Conditional posterior distributions for Gibbs sampler
  - $p(c|all) = p(c|X(t - \Delta), \Delta, y(t))$
  - $p(X(t), X(t - \Delta), \theta, \Delta | x(t), y(t), c)$
    \[= p(X(t - \Delta)|X(t), \theta, \Delta, x(t), y(t), c) \cdot p(\Delta|X(t), \theta, x(t), y(t), c) \cdot p(X(t), \theta|x(t), y(t), c)\]
    \[= p(X(t - \Delta)|X(t), \theta, \Delta, y(t), c) \cdot p(\Delta|\theta, x(t), y(t), c) \cdot p(X(t), \theta|x(t))\]

- Obtaining good posterior samples of one light curve, $(X(t), \theta|x(t))$, is a key to the successful Gibbs sampler.

- Two possible ways to sample $(X(t), \theta|x(t))$: Kelly et al. or Metropolis-Hastings in Gibbs sampler
Two Possible Samplers for One Light Curve

- Kelly et al. (2009) introduces $p(\theta | x(t))$ with $X(t)$ integrated out.
  
  $p(x(t) | X(t)) \cdot p(X(t) | \theta) \cdot p(\theta) \propto p(X(t), \theta | x(t))$
  
  $= p(X(t) | \theta, x(t)) \cdot p(\theta | x(t))$

- Alternatively we can use Metropolis-Hastings in Gibbs sampler, iteratively sampling $X(t)$ and $\theta$ from $p(X(t) | \theta, x(t))$ and $p(\theta | X(t), x(t))$ respectively.

- Comparison: 3,000 posterior samples of $\theta$ after 3,000 warming-up.

<table>
<thead>
<tr>
<th></th>
<th>median ($\mu, \sigma, \tau$)</th>
<th>sd ($\mu, \sigma, \tau$)</th>
<th>accept.rate</th>
<th>time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kelly et al.</td>
<td>(0.158, 0.0052, 358)</td>
<td>(0.11, 0.0007, 3246)</td>
<td>(0.33, 0.33, 0.35)</td>
<td>47.3</td>
</tr>
<tr>
<td>MH in Gibbs</td>
<td>(0.154, 0.0057, 290)</td>
<td>(0.10, 0.0008, 2271)</td>
<td>(NA, NA, 0.34)</td>
<td>20.9</td>
</tr>
</tbody>
</table>
TWO POSSIBLE SAMPLERS FOR ONE LIGHT CURVE

Kelly et al.

MH in Gibbs
Example 1: Data from Burud et al. (2002)

- 57 observations for each light curve.

\[ \text{R-band light curves of SBS1520+530} \]

- Their time delay estimate is \( 128 \pm 3(1\sigma) \) using \( \chi^2 \) minimization, and \( 130 \pm 3(1\sigma) \) using their iterative version of \( \chi^2 \) minimization.

- The posterior mean (median) of the time delay estimate was \( 126.8(126.5) \pm 2.1 \).
Example 1: Data from Burud et al. (2002)

- 5 chains each of which has 3,000 samples with 3,000 warming-up
- 270 seconds in total.
- Initial values
  - $\Delta: (75, 100, 125, 150, 175)$
  - $\mu, \sigma, \tau, X(t), X(t - \Delta), c: (1, 0.005, 300, x(t), y(t) - 0.7, 0.7)$
  - $\tau \sim InvGam(1, 1)$ and $\Delta \sim InvGam(1, 1)$
- Gelman-Rubin $\hat{R} = (1, 1, 1, 1, 1)$ for $(\Delta, c, \mu, \sigma, \tau)$
- Diagnosis plots for $(\Delta, c, \mu, \sigma, \tau)$
Example 1: Data from Burud et al. (2002)

For your reference, Burud et al. used $\hat{c} = 0.69$ arbitrarily to overlap red and blue points in their paper.
Example 2: Data from Kochanek et al. (2006)

- 147 observations for each light curve with wide seasonal gap.

![Light curves of HE 0435 - 1223]

- Their time delay estimate is $14.37^{+0.75}_{-0.85}$ using adjusted $\chi^2$ minimization.

- The posterior mean (median) of the time delay estimate was $17.47(17.52) \pm 0.48$. 
Example 2: Data from Kochanek et al. (2006)

- 5 chains each of which has 3,000 samples with 3,000 warming-up
- 650 seconds in total.
- Initial values
  - $\Delta$: (5, 10, 15, 20, 25)
  - $\mu, \sigma, \tau, X(t), X(t - \Delta), c : (2, 0.01, 100, x(t), y(t) - 0.78, 0.78)$
  - $\tau \sim InvGam(1, 1)$ and $\Delta \sim InvGam(1, 1)$
- Gelman-Rubin $\hat{R} = (1, 1, 1, 1, 1)$ for $(\Delta, c, \mu, \sigma, \tau)$
- Diagnosis plots for $(\Delta, c, \mu, \sigma, \tau)$
Example 2: Data from Kochanek et al. (2006)

For your reference, Kochanek et al. did not provide information on \( \hat{c} \), though they shifted it in their paper. So I arbitrarily shifted blue dots by 0.76 in \( y \)-axis on the right plot.
Discussion

- Prior choice for $\Delta$
- Sensitivity analysis for $\tau \sim InvGam(1, 1)$ and $\Delta \sim InvGam(1, 1)$
- Participation in Time Delay Challenge, an on-going blind competition


