Parametric Bayesian Approach to Time Delay Estimation

Hyungsuk Tak

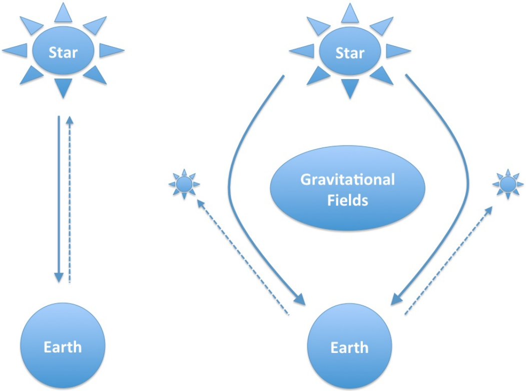
Stat 310

1 April 2014

Outline

- 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

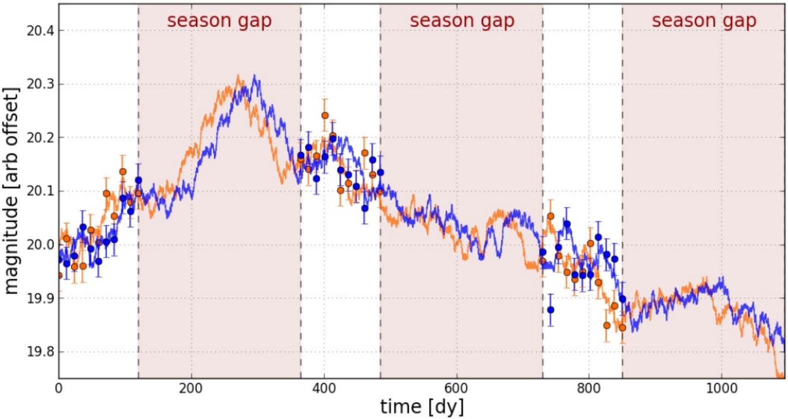
Introduction



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.

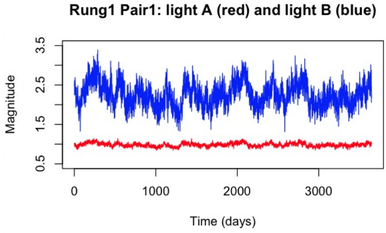
- TDC0, called a ladder, consists of 7 rungs (increasing difficulty).

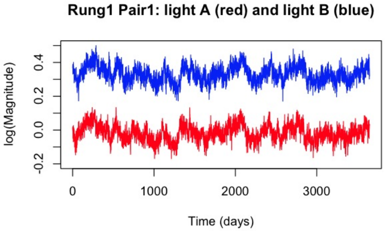
- Each rung (subscript j) has 8 data sets (subscript i).
- Each data set contains a pair of light curves with measurement errors.
- Good team's job is to estimate the time delays in each dataset, $\hat{\Delta}_{ij}$, where $i = 1, 2, \ldots, 8$ and $j = 1, 2, \ldots, 7$.

Data Description

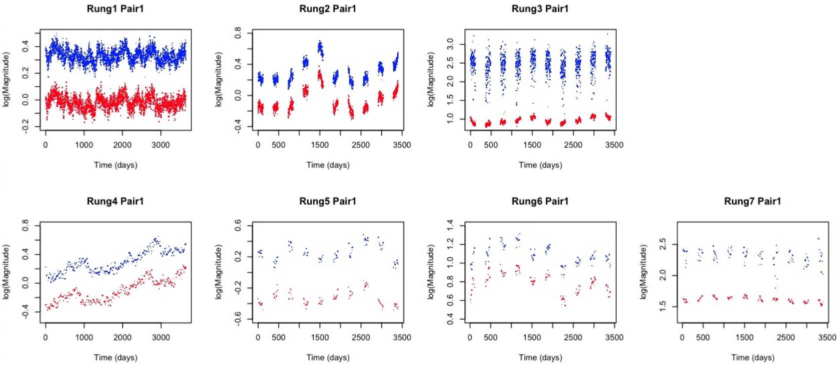
5 variables in each dataset

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





Complexities and Challenges



- Rung1 $\rightarrow$ Rung2: Seasonal gaps
- Rung2 $\rightarrow$ Rung3: More variations in the following blue light curve
- Rung1 $\rightarrow$ Rung4: Sparse (irregular) observations (sampling time)
- Rung4 $\rightarrow$ Rung5: Seasonal gaps
- Rung5 $\rightarrow$ Rung6: Non-interger time (sampling time on real line)
- Rung6 $\rightarrow$ Rung7: More variations in the following blue light curve

Popular Estimation Methods

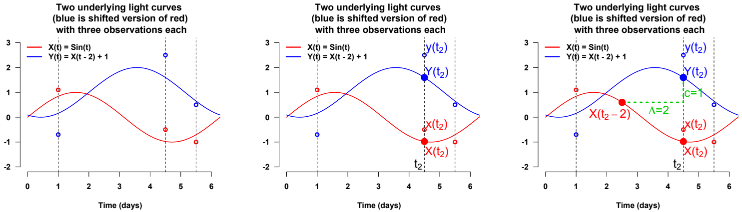
- Smoothing and χ^2-minimization (Fassnacht, 1999)
  - Smooth both light curves
  - Scale (by σ) and shift (by Δ) one smooth light curve
  - Calculate $\chi^2_{\sigma,\Delta}$ statistic
  - Find Δ minimizing $\chi^2_{\sigma,\Delta}$ on the two-dimensional grids of σ and Δ
- Smoothing and Cross-correlation (Fassnacht, 1999)
  - Smooth both light curves
  - Shift (by Δ) one smooth light curve
  - Calculate r_Δ, sample cross-correlation functions
  - Find Δ that maximizes r_Δ on the grid of Δ

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 ($D^2_{c,\Delta}$), defined as the weighted sum of squared differences of two adjoining points of the composite curve.
  - Find Δ minimizing $D^2_{c,\Delta}$ on the two-dimensional grids of c and Δ.
- 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 Δ that minimizes the weighted average variation of difference curve, $GP1(t) - GP2(t+\Delta)$, on grid of Δ.

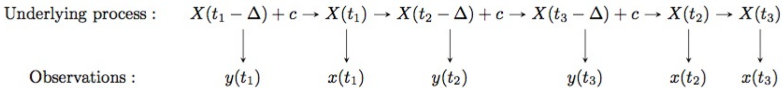
Idea and Model Specification

- $\exists$ 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:



  - 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



- Likelihood:

$$\begin{cases} x(t_j) = X(t_j) + \epsilon_j, \quad \epsilon_j \sim N(0, \delta_j^2), \\ y(t_j) = X(t_j - \Delta) + c + e_j, \quad e_j \sim N(0, \eta_j^2), \quad j = 1, 2, \ldots, n \end{cases}$$

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

Prior: Ornstein-Uhlenbeck Process

- Need to build a model for underlying (latent) light curve

  $p(X(\mathbf{t}), X(\mathbf{t}-\Delta)|\theta, \Delta) = p(X(\mathbf{t}')|\theta, \Delta)$

  - 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

- $dX(t) = -\frac{1}{\tau}\big(X(t) - \mu\big)dt + \sigma dB(t)$, where

- τ is a relaxation time, μ and σ 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

  $$X(t_j)|X(t_{j-1}), \mu, \sigma^2, \tau \sim N\Big[\text{mean: } \mu + e^{-(t_j - t_{j-1})/\tau}\big(X(t_{j-1}) - \mu\big), \text{ variance: } \frac{\tau\sigma^2}{2}\big(1 - e^{-2(t_j - t_{j-1})/\tau}\big)\Big]$$

- $p(X(\mathbf{t}')|\theta, \Delta) = p(X(t_1')|\theta, \Delta)\prod_{j=2}^{2n} p(X(t_j')|X(t_{j-1}'), \theta, \Delta)$

Hyper-prior Distribution

- 5 hyper-parameters:
  - μ is a mean parameter of underlying process
  - σ is a scale parameter of underlying process
  - τ is a relaxation time of the underlying process
  - c is a shift in y-axis
  - Δ is a shift in x-axis (time delay)
- Naively informative: $p(\theta, c, \Delta) \equiv p(\mu, \sigma^2, \tau, c, \Delta) \propto \frac{1}{\sigma} \frac{e^{-\epsilon_1/\tau}}{\tau^{\epsilon_1+1}} \frac{e^{-\epsilon_2/\Delta}}{\Delta^{\epsilon_2+1}}$
- $\tau \sim \mathit{InvGam}(\epsilon_1, \epsilon_1)$ and $\Delta \sim \mathit{InvGam}(\epsilon_2, \epsilon_2)$
- In general, a diffuse hyper-prior distribution (possibly Normal) on Δ, if we do not know which light curve is preceding

Full Posterior Distribution

- Full Posterior: $p(X(\mathbf{t}), X(\mathbf{t}-\Delta), \theta, c, \Delta | x(\mathbf{t}), y(\mathbf{t}))$

$$\propto p(x(\mathbf{t})|X(\mathbf{t}))\cdot p(y(\mathbf{t})|X(\mathbf{t}-\Delta)+c, c, \Delta) \quad \text{Likelihood}$$
$$\cdot p(X(\mathbf{t}), X(\mathbf{t}-\Delta)|\theta, \Delta) \quad \text{Prior}$$
$$\cdot p(\theta, c, \Delta) \quad \text{Hyper-prior}$$

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

Conditional Posterior Distributions

- Conditional posterior distributions for Gibbs sampler
  - $p(c|all) = p(c|X(\mathbf{t}-\Delta), \Delta, y(\mathbf{t}))$
  - $p(X(\mathbf{t}), X(\mathbf{t}-\Delta), \theta, \Delta | x(\mathbf{t}), y(\mathbf{t}), c)$

    $= p(X(\mathbf{t}-\Delta)|X(\mathbf{t}), \theta, \Delta, x(\mathbf{t}), y(\mathbf{t}), c)$

    $\cdot p(\Delta|X(\mathbf{t}), \theta, x(\mathbf{t}), y(\mathbf{t}), c) \cdot p(X(\mathbf{t}), \theta | x(\mathbf{t}), y(\mathbf{t}), c)$

    $= p(X(\mathbf{t}-\Delta)|X(\mathbf{t}), \theta, \Delta, y(\mathbf{t}), c)$

    $\cdot p(\Delta|\theta, x(\mathbf{t}), y(\mathbf{t}), c) \cdot p(X(\mathbf{t}), \theta | x(\mathbf{t}))$
- Obtaining good posterior samples of one light curve, $(X(\mathbf{t}), \theta | x(\mathbf{t}))$, is a key to the successful Gibbs sampler.
- Two possible ways to sample $(X(\mathbf{t}), \theta | x(\mathbf{t}))$: Kelly et al. or Metropolis-Hastings in Gibbs sampler

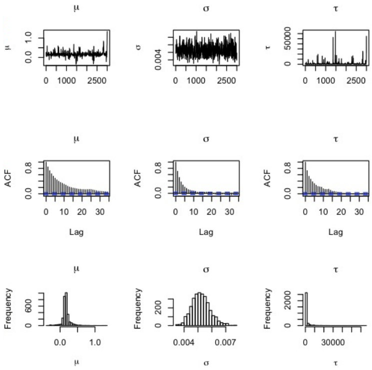
Two Possible Samplers for One Light Curve

- Kelly et al. (2009) introduces $p(\theta|x(\mathbf{t}))$ with $X(\mathbf{t})$ integrated out.
  - $$p(x(\mathbf{t})|X(\mathbf{t})) \cdot p(X(\mathbf{t})|\theta) \cdot p(\theta) \propto p(X(\mathbf{t}), \theta|x(\mathbf{t})) = p(X(\mathbf{t})|\theta, x(\mathbf{t})) \cdot p(\theta|x(\mathbf{t}))$$
- Alternatively we can use Metropolis-Hastings in Gibbs sampler, iteratively sampling $X(\mathbf{t})$ and θ from $p(X(\mathbf{t})|\theta, x(\mathbf{t}))$ and $p(\theta|X(\mathbf{t}), x(\mathbf{t}))$ respectively.
- Comparison: 3,000 posterior samples of θ after 3,000 warming-up.

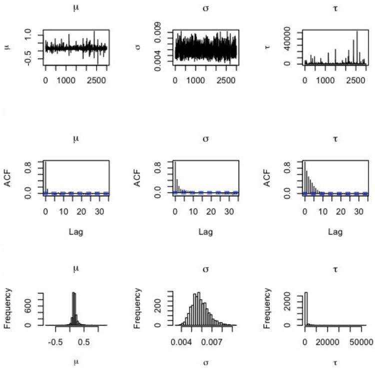
	median (μ, σ, τ)	sd (μ, σ, τ)	accept.rate	time (sec)
Kelly et al.	(0.158, 0.0052, 358)	(0.11, 0.0007, 3246)	(0.33, 0.33, 0.35)	47.3
MH in Gibbs	(0.154, 0.0057, 290)	(0.10, 0.0008, 2271)	(NA, NA, 0.34)	20.9

Two Possible Samplers for One Light Curve

Kelly et al.

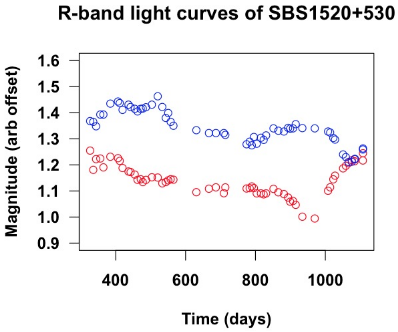


MH in Gibbs



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

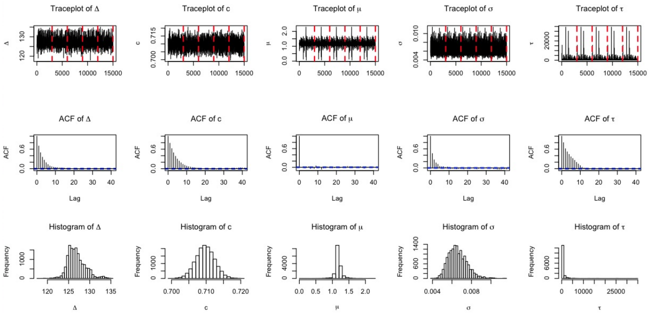
- 57 observations for each light curve.



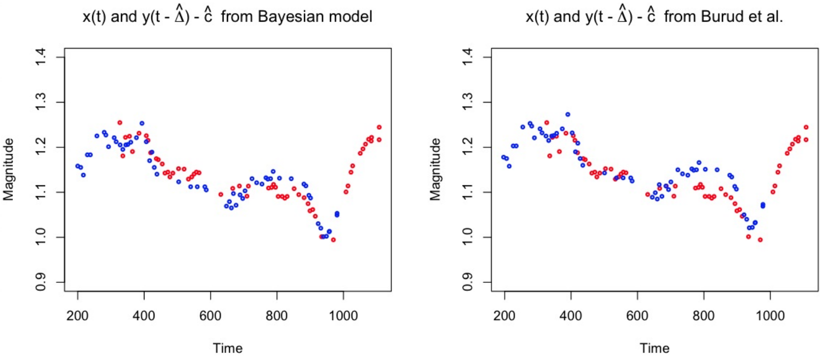
- Their time delay estimate is $128 \pm 3(1\sigma)$ using χ^2 minimization, and $130 \pm 3(1\sigma)$ using their iterative version of χ^2 minimization.
- The posterior mean (median) of the time delay estimate was $126.8(126.5) \pm 2.1$.

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
  - Δ: (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 $(\Delta, c, \mu, \sigma, \tau)$



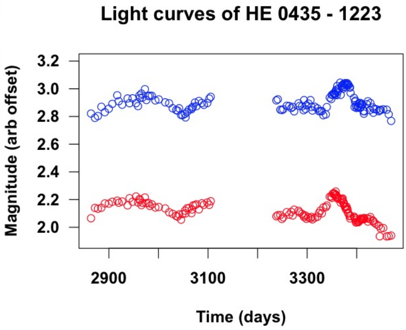
Example 1: Data from Burud et al. (2002)



- 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)

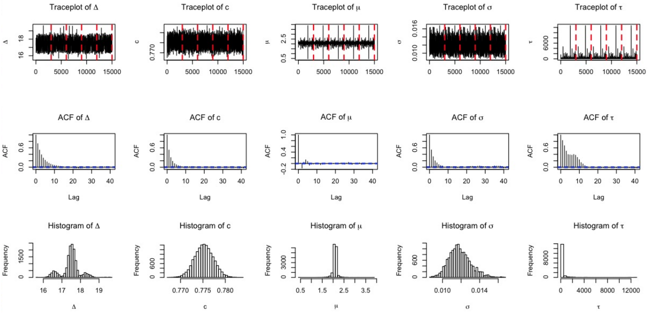
- 147 observations for each light curve with wide seasonal gap.



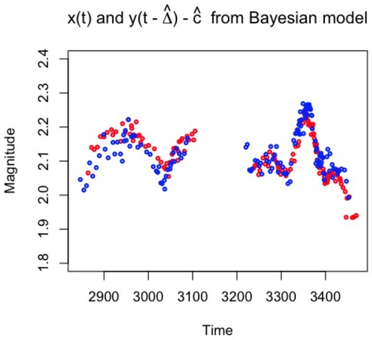
- Their time delay estimate is $14.37^{+0.75}_{-0.85}$ using adjusted χ^2 minimization.
- The posterior mean (median) of the time delay estimate was $17.47(17.52) \pm 0.48$.

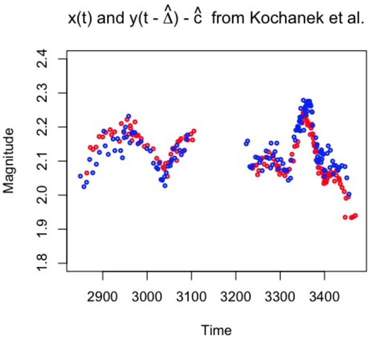
Example 2: Data from Kochanek et al. (2006)

- 5 chains each of which has 3,000 samples with 3,000 warming-up
- 650 seconds in total.
- Initial values
  - Δ: (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 Δ
- Sensitivity analysis for $\tau \sim \mathit{InvGam}(1,1)$ and $\Delta \sim \mathit{InvGam}(1,1)$
- Participation in Time Delay Challenge, an on-going blind competition

Reference

1. I. Burud, J. Hjorth, F. Courbin, J. Cohen, P. Magain, A. Jaunsen, A. Kaas, C. Faure, and G. Letawe (2004) "Time delay and lens redshift for the doubly imaged BAL quasar SBS 1520+530" *Astronomy & Astrophysics manuscript no.*
2. G. Dobler, C. Fassnacht, T. Treu, P. Marchall, K. Liao, A. Hojjati, E. Linder, and N. Rumbaugh (2013) "Strong lens time delay challenge: I. Experimental design."
3. C. Fassnacht, T. Pearson, A. Readhead, I. Browne, L. Koopmans, S. Myers, and P. Wilkinson (1999) "A determination of H_0 with the class graviatational lens B1608+656. I. time delay measurements with the VLA" *The Astrophysical Journal*, **527**, 498-512.
4. A. Hojjati, A. Kim, and E. Linder (2013) "Robust Strong Lensing Time Delay Estimation" *in progress*.
5. B. Kelly, J. Bechtold, and A. Siemiginowska (2009) "Are the variations in quasar optical flux driven by thermal fluctuation?" *The Astrophysical Journal*, **698**, 895 - 910.
6. C. Kochanek, N. Morgan, E. Falco, B. McLeod, and J. Winn (2006) "The time delays of gravitational lens HE 0435-1223: An early-type galaxy with a rising rotation curve" *The Astrophysical Journal*, **640**, 47-61.
7. M. Tewes, F. Courbin, and G. Meylan (2013) "COSMOGRAIL: the COSmological MOnitoring of GRAvitational Lenses" *Astronomy & Astrophysics manuscript no.*