

BAYESIAN ESTIMATION OF TIME DELAYS BETWEEN UNEVENLY SAMPLED SIGNALS

Markus Harva

Adaptive Informatics Research Centre
Helsinki University of Technology
P.O. Box 5400, FI-02015 TKK, Espoo, Finland
markus.harva@hut.fi

Somak Raychaudhury

School of Physics and Astronomy
University of Birmingham
Birmingham B15 2TT, UK
somak@star.sr.bham.ac.uk

ABSTRACT

A method for estimating time delays between signals that are irregularly sampled is presented. The approach is based on postulating a latent variable model from which the observed signals have been generated and computing the posterior distribution of the delay. This is achieved partly by exact marginalisation and partly by using MCMC methods. Experiments with artificial data show the effectiveness of the proposed approach while results with real-world gravitational lens data provide the main motivation for this work.

1. INTRODUCTION

The estimation of a delay between two signals is concerned with the following question: given two time series $x_1(t)$ and $x_2(t)$, for which time lag τ the relation $x_1(t) = x_2(t - \tau)$ holds most accurately? The question is of importance in many fields of signal processing, from underwater acoustics to satellite positioning. Much of the work done in answering the question has dealt with evenly sampled signals and the solution is most often based on the cross correlation function or a variation of it.

The problem drastically changes when the signals are unevenly sampled. Such a situation rises frequently in certain fields, such as astronomy, where the sampling times are largely determined by outside factors, like observing conditions and scheduling. In the uneven sampling case, the cross correlation function cannot usually be exactly evaluated for any case other than $\tau = 0$. For other values, some form of interpolation needs to be performed, making the implicit assumption that the signals are slowly varying. Even if that can be assumed, it has been argued, that in noisy environments, interpolation introduces a lot of new data which has very little to do with reality. Consequently, several methods have been proposed as alternatives to the standard cross correlation function. Of these, perhaps the most widely used are the discrete correlation function [1] and especially the locally normalised version [2] of it.

A recent study [3], involving one the authors of this pa-

per, shows that the delay estimation problem with irregular sampling, mainly in the context of the determination of time lags in gravitationally lensed multiple images of distant celestial objects, is far from completely solved.

In this paper, we propose a Bayesian approach to solving the delay estimation problem with unevenly sampled signals. We formulate a latent variable model exploiting the assumption that the signals are slowly varying. We avoid any kind of interpolation on the noisy data. We derive an inference algorithm for our model partly based on exact marginalisation and partly on MCMC methods. Although the main motivation of this work is solving the delays in gravitational lensing, no domain specific assumptions are made and the method is of general applicability as long as the assumption of slow variability can be made.

In the next section we lay out the exact modelling assumptions that we make. In Section 3 the algorithm for computing the posterior probability distribution of the delay is derived. Section 4 reports both comparison experiments against other methods as well as results with real-world gravitational lensing data.

2. MODELLING ASSUMPTIONS

In the basic setting we have two sets of observations, $x_1(t)$ and $x_2(t)$, measured at time instants $t'_i, i = 1, \dots, N$. Both of these observations are due to a common source $s(t)$ but the second observation is delayed by τ . Now, with a fixed delay τ , we can think of having observations at $2N$ distinct time instants, namely at t'_i and $t'_i - \tau$. By ordering these we obtain a signal $x(t_i)$, measured at time instants $t_i, i = 1, \dots, 2N$, which incorporates the samples from both x_1 and x_2 . We denote the correspondence with $k(i) \in \{1, 2\}$ such that $x(t_i) = x_{k(i)}(t'_j)$ for some j . The source can also be shifted and scaled individually for each of the two observations. Denoting the data as $\mathbf{X} = \{x(t_i)|i\}$, the source as $\mathbf{S} = \{s(t_i)|i\}$, the scaling coefficients and the shifts as a_k and b_k , respectively, and finally the log-variances of the noise as v_k , we can write the likelihood for the model pa-

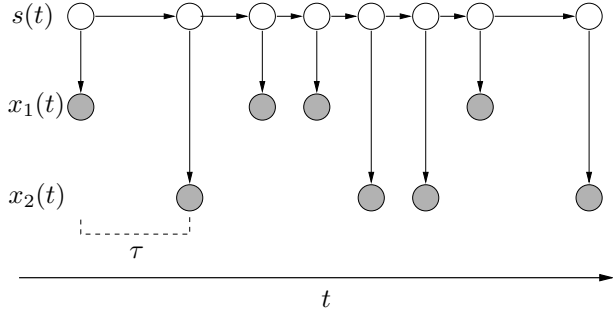


Fig. 1. The time delay τ affects the structure of the model.

rameters as

$$p(\mathbf{X}|\mathbf{S}, \{a_k\}, \{b_k\}, \{v_k\}, \tau) = \prod_{i=1}^{2N} \mathcal{N}(x(t_i) | a_{k(i)}s(t_i) + b_{k(i)}, e^{v_{k(i)}}). \quad (1)$$

Here, $\mathcal{N}(x|\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 .

The mere likelihood above doesn't relate the individual observations to each other in any way and hence using it alone it is not possible to estimate the delay. The additional assumption we make is that the source varies slowly in time, such that the next state $s(t_{i+1})$ is dependent on the current state $s(t_i)$. We incorporate this assumption to the model by putting the following prior on \mathbf{S} :

$$p(\mathbf{S}|w, \gamma, \tau) = \prod_{i=2}^{2N} p(s(t_i)|s(t_{i-1}), w, \gamma) \times p(s(t_1)) = \prod_{i=2}^{2N} \mathcal{N}(s(t_i)|s(t_{i-1}), (t_i - t_{i-1})^\gamma e^w) \times \mathcal{N}(s(t_1)|0, \sigma_s^2). \quad (2)$$

The prior states that the deviation of the source from the previous time instant is proportional to the temporal distance between the current instant and the previous instant. We see this formulation as least restrictive while still incorporating the prior knowledge. The constant σ_s^2 specifies the prior variance of the source at the first time instant. Since the data can be normalised as a preprocessing step, we can use a moderate value of $\sigma_s^2 = 10^2$.

On the surface, it might seem that the delay has no role in Equations (1) and (2). It of course does, but the effect is of structural nature as it affects the combined time instants t_i and consequently the order in which the data samples appear in the model. This is illustrated in Figure 1.

For the parameters we choose the following priors:

$$\begin{aligned} p(\tau) &= \mathcal{N}(\tau | \mu_\tau, \sigma_\tau^2), \\ p(w) &= \mathcal{N}(w | -5, 5^2), \\ p(v_k) &= \mathcal{N}(v_k | -5, 5^2). \end{aligned} \quad (3)$$

The constants μ_τ and σ_τ^2 are chosen to reflect the earlier knowledge of the delay (if such exists), but so that they do not constrain the delay too much. We have considered γ both as a variable and a constant, but found that little difference is made in estimating it, and so we keep it fixed to $\gamma = 2$.

For the scale and shift parameters to be identifiable, we fix a_1 and b_1 to one and zero, respectively. Because of the normalisation, the values of the parameters a_2 and b_2 should not be too far from a_1 and b_1 . Hence, for them we have the priors

$$\begin{aligned} p(a_2) &= \mathcal{N}(a_2 | 1, 1), \\ p(b_2) &= \mathcal{N}(b_2 | 0, 1). \end{aligned}$$

When we have knowledge of the uncertainties related to the measuring process, we want to incorporate that to the model as well. In that case, we actually postulate \mathbf{X} as hidden variables from which we have the noisy observations $\mathbf{Y} = \{y(t_i)|i\}$ with known standard deviations $\sigma_y(t_i)$:

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{2N} \mathcal{N}(y(t_i)|x(t_i), \sigma_y^2(t_i)). \quad (4)$$

This concludes the formulation of the model in the case of two observed signals. The same construction extends straightforwardly to the multiple (> 2) signals case. Also, when several measurements from different wavelengths or distinct time intervals are available, we can use them jointly. If we have M datasets, we also have M set of parameters, excluding the delay (delays) which is (are) common to all datasets.

3. LEARNING THE MODEL

We are only interested in the posterior distribution of the delay $p(\tau|\mathbf{Y})$ and consider all the other variables in the model to be nuisance parameters which we would like to marginalise out. In the next subsection we show how to get rid of most of the nuisance parameters by exact marginalisation which results in a recursive formula for the marginal likelihood. This reduces the dimensionality of the parameter space radically making sampling from the rest of the variables effective. The sampling procedure is discussed in Subsection 3.2.

3.1. Marginalisation

We denote by $\boldsymbol{\theta}$ all the time-independent parameters in the model, excluding the delay. We are going to compute the marginal likelihood $p(\mathbf{Y}|\tau, \boldsymbol{\theta})$. We can easily integrate out the error model:

$$\begin{aligned} p(\mathbf{Y}|\mathbf{S}, \tau, \boldsymbol{\theta}) &= \int d\mathbf{X} p(\mathbf{Y}|\mathbf{X}) p(\mathbf{X}|\mathbf{S}, \tau, \boldsymbol{\theta}) \\ &= \int d\mathbf{X} \left\{ \prod_{i=1}^{2N} p(y(t_i)|x(t_i)) p(x(t_i)|s(t_i), \tau, \boldsymbol{\theta}) \right\} \\ &= \prod_{i=1}^{2N} \int dx(t_i) \mathcal{N}(y(t_i)|x(t_i), \sigma_y^2(t_i)) \\ &\quad \times \mathcal{N}(x(t_i)|a_{k(i)}s(t_i) + b_{k(i)}, e^{v_{k(i)}}) \\ &= \prod_{i=1}^{2N} \frac{1}{|a_{k(i)}|} \mathcal{N}(\hat{y}(t_i)|s(t_i), \tilde{y}(t_i)), \end{aligned}$$

where

$$\hat{y}(t_i) = \frac{y(t_i) - b_{k(i)}}{a_{k(i)}} \quad \text{and} \quad \tilde{y}(t_i) = \frac{e^{v_{k(i)}} + \sigma_y^2(t_i)}{a_{k(i)}^2}.$$

To marginalise over the source, we need to compute the integral

$$\begin{aligned} p(\mathbf{Y}|\tau, \boldsymbol{\theta}) &= \int d\mathbf{S} p(\mathbf{Y}|\mathbf{S}, \tau, \boldsymbol{\theta}) p(\mathbf{S}|\tau, \boldsymbol{\theta}) \\ &= \int ds(t_1) \cdots ds(t_{2N}) \left\{ \prod_{i=2}^{2N} \frac{1}{|a_{k(i)}|} \mathcal{N}(\hat{y}(t_i)|s(t_i), \tilde{y}(t_i)) \right. \\ &\quad \times \mathcal{N}(s(t_i)|s(t_{i-1}), (t_i - t_{i-1})^\gamma e^w) \left. \right\} \\ &\quad \times \frac{1}{|a_{k(1)}|} \mathcal{N}(\hat{y}(t_1)|s(t_1), \tilde{y}(t_1)) \mathcal{N}(s(t_1)|0, \sigma_s^2). \end{aligned}$$

Making use of Gaussian identities, this can be evaluated by integrating $s(t_i)$ out one by one starting from $s(t_1)$ and the procedure yields the expression

$$\begin{aligned} p(\mathbf{Y}|\tau, \boldsymbol{\theta}) &= \left\{ \prod_{i=2}^{2N} \frac{1}{|a_{k(i)}|} \right. \\ &\quad \times \mathcal{N}(\hat{y}(t_i)|\hat{s}(t_{i-1}), (t_i - t_{i-1})^\gamma e^w + \tilde{s}(t_{i-1}) + \tilde{y}(t_i)) \left. \right\} \\ &\quad \times \frac{1}{|a_{k(1)}|} \mathcal{N}(\hat{y}(t_1)|0, \sigma_s^2 + \tilde{y}(t_1)), \end{aligned}$$

where the quantities \hat{s} and \tilde{s} are computed from the recursive formulae

$$\begin{aligned} \tilde{s}(t_1) &= (\sigma_s^{-2} + \tilde{y}^{-1}(t_1))^{-1} \\ \hat{s}(t_1) &= \tilde{s}(t_1) \hat{y}(t_1) / \tilde{y}(t_1) \end{aligned}$$

and for $i > 1$

$$\begin{aligned} \tilde{s}(t_i) &= \left\{ [(t_i - t_{i-1})^\gamma e^w + \tilde{s}(t_{i-1})]^{-1} + \tilde{y}^{-1}(t_i) \right\}^{-1} \\ \hat{s}(t_i) &= \tilde{s}(t_i) \left\{ \hat{s}(t_{i-1}) / [(t_i - t_{i-1})^\gamma e^w + \tilde{s}(t_{i-1})] \right. \\ &\quad \left. + \hat{y}(t_i) / \tilde{y}(t_i) \right\}. \end{aligned}$$

3.2. Sampling procedure

The final step in obtaining summaries from $p(\tau|\mathbf{Y})$ is to sample from $p(\tau, \boldsymbol{\theta}|\mathbf{Y})$. As an alternative, we considered numerically marginalising over $\boldsymbol{\theta}$, but the dimensionality is still high enough, especially in the case of more than one delay and more than two datasets, to render the task intractable.

The Metropolis algorithm (see e.g. [4]) is particularly suitable here, as it only requires us to be able to evaluate the unnormalised posterior density; something we can readily accomplish by multiplying the marginal likelihood of the previous section with the prior. In addition, a suitable jumping distribution needs to be formulated such that the resulting Markov Chain converges reasonably fast.

A crucial part in using MCMC methods is the evaluation of the convergence of the chain. If the Markov chain has not converged to its equilibrium distribution, the samples do not come from the desired distribution making further inferences invalid. Although there is no fool proof method to assess the convergence, there exists several schemes that have been proved to be useful in practise. One such method, and the one we have used here, is to compute the so-called potential scale reduction factor [5] from several parallel Markov chains started from random initial states. It should give an estimate of how much the scale of the empirical posterior could be reduced by obtaining more samples. Practically, it compares the first and second order moments of the parallel chains.

4. EXPERIMENTAL RESULTS

The main motivation of this work is the estimation of delays in gravitational lensing systems. However, comparing methods using real datasets is very difficult as the true delay remains unknown. Hence, we have made comparison experiments using artificial datasets, as then we can make well justified claims about the accuracy of different methods. These are reported in the next subsection. Results with real-world gravitational lensing datasets are reported and discussed in Section 4.2.

4.1. Controlled comparisons

We compare the proposed approach to the following three widely used methods: discrete correlation function (DCF) [1],

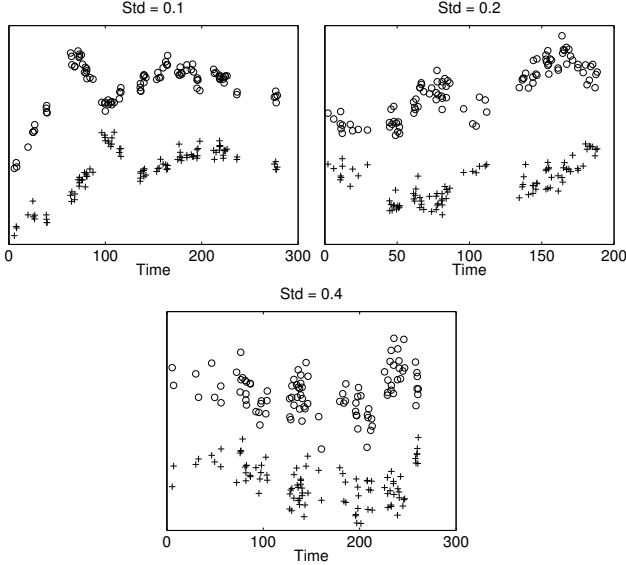


Fig. 2. Examples of artificial datasets with various noise levels.

locally normalised discrete correlation function (LNDCF) [2], and linear interpolation followed by standard cross correlation (similar scheme used e.g. in [6]).

We did repeated experiments with artificial datasets generated from our model with a fixed delay of 35 units. The datasets were comprised of two observed signals having one hundred samples each. The sampling time instants t'_i where generated from the following mixture model:

$$\Delta t \sim 0.8 \times \mathcal{E}(1) + 0.2 \times \mathcal{E}(0.1)$$

$$t'_i = t'_{i-1} + \Delta t,$$

where $\mathcal{E}(\lambda)$ denotes the exponential distribution with inverse scale λ . This model was used to simulate the gaps that are typical in astronomical datasets. The source was generated using Eq. (2) with $\gamma = 2$ and $w = 2 \ln 0.05$. The observations were generated from the distribution (1) with $a_1 = 1$, $b_1 = 0$, $a_2 = 0.8$ and $b_2 = 0.2$. The noise variances e^{v_k} were same for both observations and three different levels (resulting in three sets of datasets) of it was used, namely 0.1^2 , 0.2^2 and 0.4^2 . Each of the three groups contained 225 datasets and examples from each of them are shown in Figure 2.

The DCF and LNDCF methods contain one tunable parameter: the binning size. To make the comparison as fair as possible, we optimised this by trying various binsizes between 5 and 60. We found that the binsize of 10 produced the best results on average with both of the methods. With the linear interpolation approach one can vary the sampling frequency. Again we tried several possibilities between 0.01 and 1.0, but found that between this interval, it didn't make much of a difference, so we selected the value 0.05. With

all these methods, we computed the value of the correlation function between 0 and 70 with stepping of 0.2 and selected the delay with the highest value as the estimate.

The proposed approach contains several tunable parameters in the proposal distribution. However, as opposed to the comparison methods, these can be selected without knowing the true delay, by monitoring the rejection ratio and adjusting the parameters appropriately. The prior on τ (see Eq. (3)) was chosen to be very vague with $\mu_\tau = 0$ and $\sigma_\tau = 200$. We drew 10000 samples from five independent chains having random initial value for the delay selected from the uniform distribution on $[0, 70]$. To reduce the possibility of including samples not coming from the posterior due to failure to converge, we tried to prune one to two of the chains by computing the potential scale reduction factor for each of the subset of chains and selecting the subset with the lowest value. Finally, half of the samples from each of the chains were discarded as the burn-in period, and the rest of the samples were combined. To be able to compare against the other methods, we collapsed the posterior distribution of the delay to a point estimate by computing the sample mean.

The accuracy of the estimates was measured using the average error defined as

$$AE = \frac{1}{225} \sum_{i=1}^{225} |\hat{\tau}_i - 35|,$$

where $\hat{\tau}_i$ is the estimate for dataset i . Ideally the AE would be zero, meaning an exactly correct estimate every single time. The scheme of randomly guessing a delay between 0 and 70 can be thought of as the baseline (in the experimental setting adopted, the comparison methods could produce an estimate only within this interval). Using that scheme, the average error would be (on average) 17.5.

Figure 3 shows the average errors as well as the 25%, 50% and 75% error quantiles for the different methods with the various noise levels. DCF seems to give consistently bad results almost independent of the amount of noise. LNDCF and the interpolation approach perform at a very similar level. The proposed approach is better on average with all noise levels and the difference becomes more prominent as the noise condition gets worse.

4.2. Estimating the delays in gravitational lensing systems

Gravitational lensing occurs when the light coming from a distant quasar is bent by the gravitational potential of an intermediate galaxy such that several images of the source are observed. Relativistic effects as well as the different lengths of the paths affect the time it takes for the photons originating from the source to travel to the observer. This

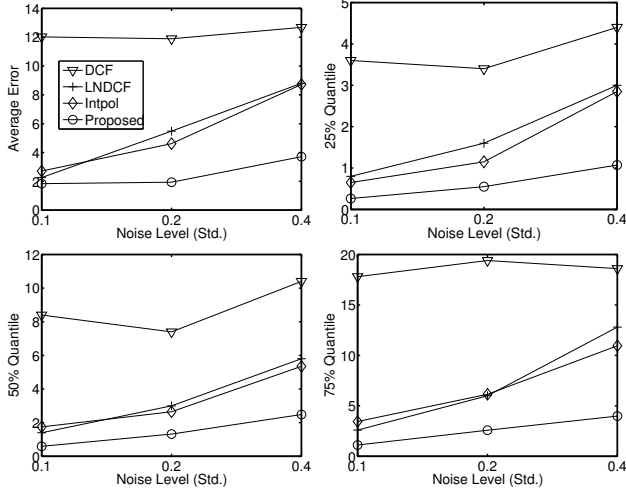


Fig. 3. Error statistics for different methods.

is perceived as a delay in the variations of the intensity profiles between the images. The significance of estimating the delays in such systems stems from the early observation that they can be used in determining important cosmological quantities [7].

We have determined the delays in several lensing systems and contrasted the results against earlier findings. The whole study is mainly interesting to the astronomical community and will be detailed elsewhere. Since the true delays are and will remain unknown, we cannot make well justified claims of having measured them better or worse than our predecessors. Instead, here we shall point out the benefits of the kind of analysis we are doing in the sense that, in addition of being rather accurate, it equips us with full probability distributions over the quantities of interest. This is very important knowledge since the uncertainties in the delays will translate to uncertainties in the derived quantities.

We shall concentrate on two systems which lie at both extremes on the well-determined / ill-determined axis. These are the systems B0218+357 (with two distinct images AB, hereafter B0218) and PG1115+080 (with at least three distinct images ABC, hereafter PG1115). The intensity profiles obtained from the systems are shown in Figure 4. In summary, Table 1 shows our results and previous measures of the delays. With B0218 we find a delay that is very well in match with the previous measure. The situation is somewhat different with PG1115. The results are not completely contradictory to the earlier findings but not in exact harmony either. Then again, with PG1115 there has been controversy before this, as is visible in the Table 1. Looking at the posterior distribution of the delays provided by our method (shown in Figure 5), we can gain some insight why this happens. The posterior over the delay in B0218

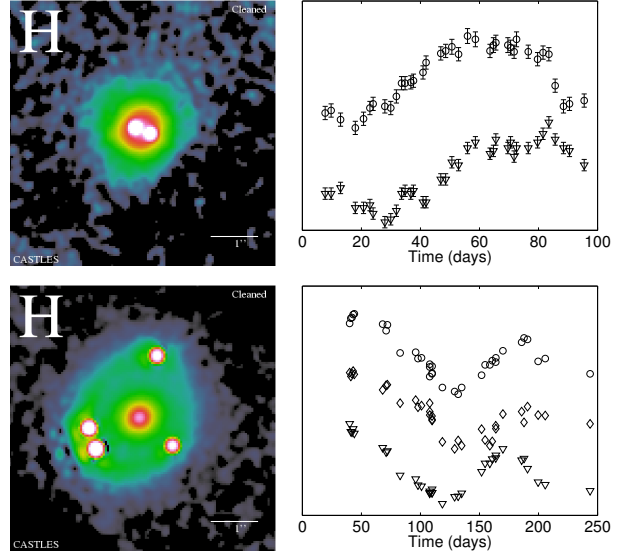


Fig. 4. Top: the two images of B0218 and the corresponding intensity measurements. Bottom: The four images of PG1115 with the light curves (the two images close to each other are merged). The images were obtained from CASTLES [8].

Table 1. Our estimates of time delays compared to previous results

System	Im.	Our Delay	Previous Measures	Ref.
B0218+357		10.9 ± 0.7	10.5 ± 0.4	[9]
PG1115+080	AC	-11.7 ± 1.7	-13 ± 1	[10]
			-9.4 ± 3.4	[11]
	BC	-22.7 ± 1.8	-25 ± 1	[10]
			-23.7 ± 3.4	[11]

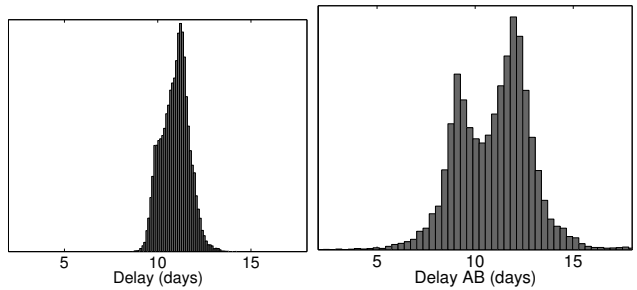


Fig. 5. Left: the posterior over the delay in B0218. Right: same for the delay AB in PG1115.

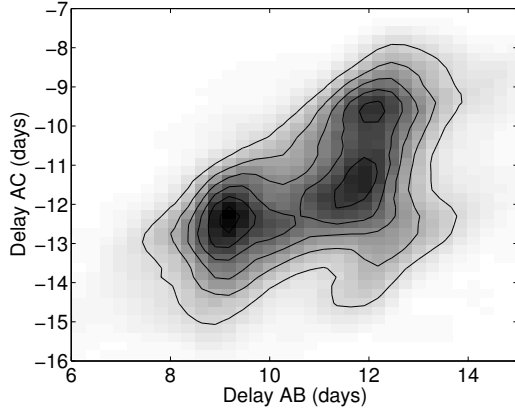


Fig. 6. The posterior distribution over the two delays in the lensing system PG1115.

is very well behaved, having the probability mass tightly concentrated around the mean value. The situation is rather different with PG1115; the posterior of one of the delays is badly multimodal spanning a wide range of values from 5 to 15. The uncertainty is even more visible in the joint distribution of the two delays shown in Figure 6. It has at least two strong modes. In this light, the posterior average and standard deviation that we customarily report in the table, do not make that much sense. But since the two strong modes have almost equal probability mass, we do not feel comfortable in computing the subsequent quantities based on just one of them either. Rather, we feel that the indeterminacy pointed out by this analysis should be taken as a hint that the system PG1115, or at least the so far observed data from it, is not the best one for the computation of the derived quantities, and any cosmological quantities derived from it should incorporate the posterior distribution in its error estimates.

5. CONCLUSIONS

The estimation of a delay between unevenly sampled signals is a recurring problem in certain fields. It is also a drastically different problem compared to its evenly sampled counterpart, since the standard cross-correlation methods are not well justified and often produce questionable results. In this paper we proposed a Bayesian approach to solve the problem, derived the learning procedure and showed its effectiveness over alternative approaches.

Acknowledgements

We wish to thank Antti Honkela for useful comments and discussion.

6. REFERENCES

- [1] R. A. Edelson and J. H. Krolik, “The discrete correlation function: a new method for analysing unevenly sampled variability data,” *The Astrophysical Journal*, vol. 333, pp. 646–659, 1988.
- [2] J. Lehar, J. N. Hewitt, B. F. Burke, and D. H. Roberts, “The radio time delay in the double quasar 0957+561,” *The Astrophysical Journal*, vol. 384, pp. 453–466, 1992.
- [3] J. C. Cuevas-Tello, P. Tiño, and S. Raychaudhury, “How accurate are the time delay estimates in gravitational lensing?,” *Astronomy & Astrophysics*, 2006, Submitted.
- [4] David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- [5] S. P. Brooks and A. Gelman, “General methods for monitoring convergence of iterative simulations,” *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998.
- [6] T. Kundic, W. N. Colley, et al., “A robust determination of the time delay in 0957+561A, B and a measurement of the global value of Hubble’s constant,” *The Astrophysical Journal*, vol. 482, pp. 75–82, 1997.
- [7] S. Refsdal, “On the possibility of determining Hubble’s parameter and the masses of galaxies from the gravitational lens effect,” *Monthly Notices of the Royal Astronomical Society*, vol. 128, pp. 307–310, 1964.
- [8] C. S. Kochanek, E. E. Falco, et al., “Website of CfA-Arizona space telescope lens survey,” <http://www.cfa.harvard.edu/castles/>.
- [9] A. D. Biggs, I. W. A. Browne, et al., “Time delay for the gravitational lens system B0218+357,” *Monthly Notices of the Royal Astronomical Society*, vol. 304, pp. 349–358, 1999.
- [10] R. Barkana, “Analysis of time delays in the gravitational lens PG 1115+080,” *The Astrophysical Journal*, vol. 489, pp. 21–28, 1997.
- [11] P. L. Schechter, “The quadruple gravitational lens PG 1115+080: Time delays and models,” *The Astrophysical Journal Letters*, vol. 475, pp. L85–L88, 1997.