

Probabilistic Modelling for Delay Estimation in Gravitationally Lensed Photon Streams

Sultanah Al Otaibi^{1,2}, Peter Tiño¹, and Somak Raychaudhury^{3,4,5}

¹ School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
P.Tino@cs.bham.ac.uk

² College of Computer and Information Sciences, King Saud University,
Riyadh 12371, Saudi Arabia

³ School of Physics and Astronomy, University of Birmingham,
Birmingham B15 2TT, UK

⁴ Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India

⁵ Department of Physics, Presidency University, Kolkata 700073, India

Abstract. We test whether a more principled treatment of delay estimation in lensed photon streams, compared with the standard kernel estimation method, can have benefits of more accurate (less biased) and/or more stable (less variance) estimation. To that end, we propose a delay estimation method in which a single latent inhomogeneous Poisson process underlying the lensed photon streams is imposed. The rate function model is formulated as a linear combination of nonlinear basis functions. Such unifying rate function is then used in delay estimation based on the corresponding Innovation Process. This method is compared with a more straightforward and less principled baseline method based on kernel estimation of the rate function. Somewhat surprisingly, the overall emerging picture is that the theoretically more principled method does not bring much practical benefit in terms of the bias/variance of the delay estimation. This is in contrast to our previous findings on daily flux data.

Keywords: Gravitational lensing · Non-homogeneous Poisson process · Kernel estimation methods

1 Introduction

Time delays between images of strongly-lensed distant variable sources can serve as a valuable tool for cosmography, provided that time delays between the image fluxes can be accurately measured (e.g. [8, 16]). A number of methods have been developed to accurately estimate time delays. These include the dispersion spectra method [3] and kernel-based method with variable width (K-V) [4, 5]. Actively studied strong quasars with time-delay measurements include RXJ1131–1231 [21] and B1608+656 [6, 8]; Q0957+561 (e.g. [9]). Available data are usually in the form of daily measurements which can be used to predict longer (days and months) delays. Current methods in astrophysics are solely

rooted in this scenario. However, when countering the problem of shorter delays (e.g. hours), daily measurements are insufficient and one needs to investigate the individual arrival times of photons.

Poisson process can be applied as a model for photon streams [15]. To resolve the delay in gravitationally lensed photon streams one can use the standard kernel based estimation of the inhomogeneous Poisson process rate function on individual photon streams and then try to time-shift the rate function estimates so as the overlap is maximized. Another, more principled alternative is to impose that the source of the delayed photon streams is the same and we simply observe different realizations from the same inhomogeneous Poisson process, gravitationally delayed in time. We study whether, comparing with the standard kernel based baseline, such a principled approach can bring benefits in terms of more stable (less variance) estimation.

Normally, delay estimation would be done over streams of photons from a given energy band and then unified over a multitude of energy bands. The baseline and principled delay estimation methods are then compared in a controlled experimental setting using synthetic photon fluxes with known imposed delay from a variety of inhomogeneous processes assumed to come from a single energy band. To our best knowledge this is the first systematic study that addresses the problem of delay estimation on lensed photon streams. We do not perform experiments on real data, since no large real photon streams from known delayed systems with short time delay are available. Nevertheless, this study serves as a proof of concept and can be readily used once appropriate lensed photon streams become available.

2 Kernel Based Delay Estimation in Lensed Photon Streams

For the sake of simplicity we will deal with the case of two lensed photon streams A and B from the same source. All techniques presented in this paper can be easily generalized to multiple streams. We assume that the observed photon streams can be accounted for by a Poisson process (e.g. [18]). In the non-homogeneous Poisson process (NHPP) the mean rate function $\lambda(s)$ varies over time s . Given a series of arrival times s_1, s_2, \dots, s_S over an interval $[0, T]$, the rate function is commonly estimated by imposing a (Gaussian) kernel of width r on top of each arrival time s_i ,

$$K_g(s; s_i, r) = \exp \left\{ -\frac{(s - s_i)^2}{2r^2} \right\}. \tag{1}$$

The rate function estimate (up to scaling) is then [12–14]

$$\hat{\lambda}(s) = \sum_{i=1}^S K_g(s; s_i, r). \tag{2}$$

We will refer to this method as Kernel Rate Estimation (KRE).

Suppose that we observe two lensed photon streams $\{s_i^A\}_{i=1}^{S^A}$ and $\{s_i^B\}_{i=1}^{S^B}$ from the same source. On each stream we produce a kernel based estimate of the rate function $\hat{\lambda}^A(s)$, $\hat{\lambda}^B(s)$. Given a suggested time delay Δ , the closeness of the rate estimates (under the delay Δ) can be evaluated e.g. through the mean square difference eventuated on a regular grid of time stamps $\{z_j\}_{j=1}^Z$ in the relevant time interval,

$$d_2(\hat{\lambda}^A, \hat{\lambda}^B; \Delta) = \frac{1}{Z} \sum_{j=1}^Z (\hat{\lambda}^A(z_j) - \hat{\lambda}^B(z_j))^2. \tag{3}$$

The delay is then estimated through minimization of $d_2(\hat{\lambda}^A, \hat{\lambda}^B; \Delta)$ with respect to Δ (e.g. via gradient descent). In the following sections we will introduce two variants of delay estimation based on innovation process corresponding to the underlying Poisson process.

3 Innovation Process Based Estimation (IPE)

Recall that if event counts can be modeled by Poisson distribution with mean rate λ , then the inter-arrival times are distributed with exponential distribution with mean λ^{-1} . We denote the differences between two consecutive arrival times by $d^A = \{d_i^A\}_{i=1}^{D^A}$ and $d^B = \{d_i^B\}_{i=1}^{D^B}$, where $d_i^A = s_{i+1}^A - s_i^A$ and $d_i^B = s_{i+1}^B - s_i^B$, respectively. Our goal is to find a probabilistic model that maximizes the probability $P(d^A, d^B | \lambda(s))$,

$$P(d^A, d^B | \lambda^A(s), \lambda^B(s)) = \prod_{i=1}^{D^A} P(d_i^A | \lambda^A(s_i; \mathbf{w})) \prod_{i=1}^{D^B} P(d_i^B | \lambda^B(s_i; \mathbf{w}, \Delta)), \tag{4}$$

where $P(d | \lambda) = \lambda e^{-\lambda d}$. We impose a kernel based model on the *common rate function* underlying both streams (expressed for stream A):

$$\lambda^A(s) = \sum_{j=1}^J w_j K_g(s; c_j, r_o) = \mathbf{w}^\top K_g(s; \mathbf{c}, r_o), \tag{5}$$

with kernels of width r_o , centered at $c_j, j = 1, 2 \dots J$ and the J free parameters w_j collected in vector \mathbf{w} . $K_g(s; \mathbf{c}, r_o)$ is a vector of kernel evaluations $K_g(s; c_j, r_o)$ at all centers of $\mathbf{c} = (c_1, c_2, \dots, c_J)$. The rate function of stream B is a time-delayed (by Δ) version of the one for stream A:

$$\lambda^B(s) = \sum_{j=1}^J w_j K_g(s; c_j - \Delta, r_o) = \mathbf{w}^\top K_g(s; \mathbf{c} - \Delta, r_o), \tag{6}$$

We thus obtain

$$P(d^A, d^B | \lambda^A(s), \lambda^B(s)) = \prod_{i=1}^{D^A} \lambda^A(s_i) e^{-\lambda^A(s_i) d_i^A} \prod_{i=1}^{D^B} \lambda^B(s_i) e^{-\lambda^B(s_i) d_i^B}, \tag{7}$$

leading to the error functional (negative log likelihood),

$$E = - \sum_{i=1}^{D^A} (\log \lambda^A(s_i) - \lambda^A(s_i) d_i^A) - \sum_{i=1}^{D^B} (\log \lambda^B(s_i) - \lambda^B(s_i) d_i^B). \quad (8)$$

We minimize E w.r.t two parameters (\mathbf{w}, Δ) via gradient descent. To that end we plug (5) and (6) into (8),

$$E = - \sum_{i=1}^{D^A} \left(\log \sum_{j=1}^J w_j K_g(s_i^A; c_j, r_o) - d_i^A \sum_{j=1}^J w_j K_g(s_i^A; c_j, r_o) \right) \\ - \sum_{i=1}^{D^B} \left(\log \sum_{j=1}^J w_j K_g(s_i^B; c_j - \Delta, r_o) - d_i^B \sum_{j=1}^J w_j K_g(s_i^B; c_j - \Delta, r_o) \right), \quad (9)$$

leading to,

$$\frac{\partial E}{\partial \mathbf{w}} = - \sum_{i=1}^{D^A} \left(\frac{K_g(s_i^A; \mathbf{c}, r_o)}{\mathbf{w}^\top K_g(s_i^A; \mathbf{c}, r_o)} - d_i^A K_g(s_i^A; \mathbf{c}, r_o) \right) \\ - \sum_{i=1}^{D^B} \left(\frac{K_g(s_i^B; \mathbf{c} - \Delta \cdot \mathbf{1}, r_o)}{\mathbf{w}^\top K_g(s_i^B; \mathbf{c} - \Delta \cdot \mathbf{1}, r_o)} - d_i^B K_g(s_i^B; \mathbf{c} - \Delta \cdot \mathbf{1}, r_o) \right), \quad (10)$$

where $\mathbf{1}$ is a vector of 1's and

$$\frac{\partial E}{\partial \Delta} = - \sum_{i=1}^{D^B} \left(\frac{1}{\sum_{j=1}^J w_j \exp\left\{-\frac{(s_i^B - (c_j - \Delta))^2}{2r_o^2}\right\}} \sum_{j=1}^J w_j \exp\left\{-\frac{(s_i^B - (c_j - \Delta))^2}{2r_o^2}\right\} \frac{-2(s_i^B - (c_j - \Delta))}{2r_o^2} \right) \\ - d_i^B \sum_{j=1}^J w_j \exp\left\{-\frac{(s_i^B - (c_j - \Delta))^2}{2r_o^2}\right\} \frac{-2(s_i^B - (c_j - \Delta))}{2r_o^2} \right). \quad (11)$$

4 Parameters Initialization

Gaussian kernels have two parameters need to be determined, in particular kernel centers $\{c_j\}_{j=1}^J$ and the kernel width r . As explained above, in KRE, kernels are centered at each photon's arrival time, whereas in IPE, the centers c_j are uniformly distributed across the time period $[0, T]$.

The kernel width determines the degree of smoothing for the underlying rate function. For KRE, we apply a method for selecting the width based on the principle of minimizing the mean integrated square error (MISE) proposed by [19]. For IPE, the kernel width r_o is optimized using cross-validation method proposed in [4, 10]. The algorithm partitions the data into 10 blocks of equal length \mathcal{L} . The i -th validation set \mathcal{V}_i , $i = 1, 2, \dots, \mathcal{L}$, is obtained by collecting the

i -th element of each block. The rest of the data is the “training set”. We then fit our models on the training set and use the validation set \mathbf{V}_i to calculate the cost function E over a range of suggested width values $r_o \in (L_{r_o}, U_{r_o})$. This procedure is repeated \mathcal{L} times for each validation set $\mathbf{V}_i, i = 1, 2 \dots \mathcal{L}$. The chosen r_o is the one yielding the smallest average cost E across the folds $i = 1, 2 \dots \mathcal{L}$.

The IPE weight vector \mathbf{w} is initialized using the rate function estimates readily provided by the KRE model. However, the rate functions obtained by KRE on streams A and B need to be scaled to represent the underlying rate of the non-homogeneous Poisson process¹. Given the KRE-estimated rate functions on streams A and B, $\hat{\lambda}^A(s), \hat{\lambda}^B(s)$, respectively, the overall KRE rate function is their average

$$\hat{\lambda}(s) = \frac{\hat{\lambda}^A(s) + \hat{\lambda}^B(s)}{2}. \tag{12}$$

The scaling factor ϑ is obtained by imposing the rate function $\lambda(s) = \vartheta \hat{\lambda}(s)$ and minimizing (9) with respect to ϑ . Denoting $\hat{\lambda}(s_i^A)d_i^A$ and $\hat{\lambda}(s_i^B)d_i^B$ by q_i^A and q_i^B , respectively, it can be shown that the minimum is obtained at

$$\vartheta = \frac{D^A + D^B}{\sum_{i=1}^{D^A} q_i^A + \sum_{i=1}^{D^B} q_i^B}. \tag{13}$$

Setting of IPE weights to match the rate function $\lambda(s)$ can then be done by imposing a regular (s_1, s_2, \dots, s_N) grid on $[0, T]$, evaluating the rate values on the grid, $\mathbf{x} = (\hat{\lambda}(s_1), \hat{\lambda}(s_2) \dots \hat{\lambda}(s_N))^T$, and solving

$$\mathbf{w} = \mathbf{K}^T \mathbf{x}, \tag{14}$$

where \mathbf{K} is an $N \times N$ matrix

$$\mathbf{K} = [K_g(s_1; \mathbf{c}, r_o), K_g(s_2; \mathbf{c}, r_o), \dots K_g(s_N; \mathbf{c}, r_o)]. \tag{15}$$

5 Experiments

To test and compare the methodologies suggested above, we performed controlled experiments on synthetic data generated from non-homogeneous Poisson processes. From each given non-homogeneous Poisson processes we generated two series A and B of arrival times, the series B was then time-shifted by a known delay Δ .

The rate functions defining non-homogeneous Poisson processes were obtained by superimposing G Gaussian functions of fixed width r_g positioned on a regular grid $\{c_g\}_{g=1}^G$ in $[0, T]$,

$$\lambda(s) = \sum_{g=1}^G w_g \cdot K_g(s; c_g, r_g), \tag{16}$$

¹ Note that for the delay detection task for which the KRE method is used, no such scaling was needed - the delay is invariant to scaling the estimated rate functions by the same factor.

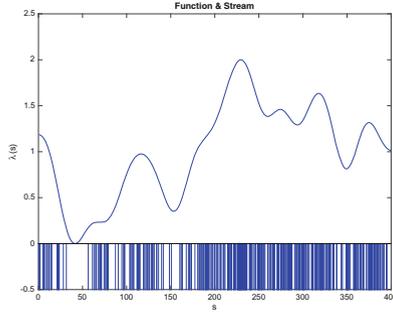


Fig. 1. An example of a test rate function and the corresponding photon stream.

where $w_g \in \mathbb{R}$ are the mixing weights generated randomly from uniform distribution on $[L_w, U_w]$. The kernel widths were set to a multiple of the kernel separation (distance between the two consecutive kernel centers) d_g , $r_g = \alpha_g \cdot d_g$. We used $T = 400$, $G = 80$, $\alpha_g = 3$, $L_w = -1$ and $U_w = 1$. The synthetic rate functions were then rescaled to the interval $[0, 2]$. Given a rate function $\lambda(s)$, the arrival times were generated using the Thinning technique [2, 7, 11, 17, 20]. An example of a test rate function and the corresponding photon stream is shown in Fig. 1.

Using this process, we generate two photon streams from the same rate function: $\{s_i^A\}$ and $\{s_j^B\}$, $i = 1, 2, \dots, S^A$ and $j = 1, 2, \dots, S^B$. To create a pair of time shifted streams, s^B is shifted in time by a delay $\Delta > 0$

$$s_i^B \leftarrow s_i^B + \Delta, \forall i = 1, 2, \dots, S^B \tag{17}$$

To prepare the streams for experiments, we cut the two streams to ensure they have the same start and end point in time.

Table 1. Statistical analysis of delay estimates. The results for each method and each imposed delay $\Delta \in \{20, 22, 25, 28\}$ are averaged over 100 test rate functions. The time delay trial values were taken from the interval $[10, 40]$ with increment of 10.

Delay	Method	$\mu \pm \sigma$	MAE	CI range	95 % CI
20	KRE	21.01±3.81	1.87	0.75	[20.26,21.76]
	IPE	20.88±3.93	1.73	0.77	[20.11,21.65]
22	KRE	22.68±4.50	3.49	0.88	[21.80,23.56]
	IPE	22.40±4.61	3.74	0.90	[21.50,23.30]
25	KRE	25.44±5.70	5.09	1.12	[24.32,26.56]
	IPE	25.55±5.96	5.54	1.17	[24.38,26.72]
28	KRE	28.56±4.30	3.27	0.84	[27.72,29.40]
	IPE	28.58±4.32	3.34	0.85	[27.73,29.43]

We performed controlled experiments, where 100 test rate functions were generated as described in Sect. 5. For each test rate function we imposed four delay values $\Delta \in \{20, 22, 25, 28\}$, resulting in 400 individual experiments. The time delay trial values were taken from the interval $[10, 40]$ with increment of 10. For each model and each imposed delay $\Delta \in \{20, 22, 25, 28\}$, we report the mean μ and standard deviation σ of the maximum-likelihood delay estimates across the set of 100 test rate functions. We also report the mean absolute error (MAE) of the delay estimates and the 95 % Credibility Interval (CI). Summary results are presented in Table 1.

6 Conclusion

In this paper, we tested whether a more principled treatment of delay estimation in lensed photon streams, compared with the standard kernel estimation method, can have benefits of a more accurate (less biased) and/or more stable (less variance) estimation. In particular, we formulated a baseline method (KRE) based on kernel estimation of the rate function of inhomogeneous Poisson process. The delay estimate is refined using gradient descent in the delay parameter on the error functional.

A more principled delay estimation relied on imposing a single latent inhomogeneous Poisson process underlying the lensed photon streams. The rate function model was formulated as a linear combination of nonlinear basis functions, thus making the non-linear model linear in the mixing parameters. We tested this idea in the Innovation Process Based Estimation (IPE).

Somewhat surprisingly, the overall emerging picture is that the theoretically more principled methods do not bring much practical benefit in terms of the bias/variance of the delay estimation. This is in contrast to our previous findings on daily flux data [1, 4, 5]. It appears that because the underlying latent rate function is represented only implicitly through the streams of arrival times weakens the stabilizing factor of the single unified intensity function that proved so useful in the case of daily flux data [1, 4, 5]. Indeed, in that case, knowing the amount of observational noise and observing noisy flux levels gave much better clues as to what the common source variability could be, thus stabilizing the delay estimation. Nevertheless, we propose that a study of the kind is useful and necessary for future developments of alternative methods for the delay estimation in lensed photon streams.

References

1. Al Otaibi, S., Tiño, P., Cuevas-Tello, J.C., Mandel, I., Raychaudhury, S.: Kernel regression estimates of time delays between gravitationally lensed fluxes. *MNRAS* **459**(1), 573–584 (2016)
2. Bratley, P., Fox, B., Schrage, L.E.: *A Guide to Simulation*, 2nd edn. Springer, New York (1987)

3. Courbin, F., Chantry, V., Revaz, Y., Sluse, D., Faure, C., Tewes, M., Eulaers, E., Koleva, M., Asfandiyarov, I., Dye, S., Magain, P., van Winckel, H., Coles, J., Saha, P., Ibrahimov, M., Meylan, G.: COSMOGRAIL: the COSMological MOonitoring of GRAvItational Lenses IX. Time delays, lens dynamics and baryonic fraction in HE 0435–1223. *Astron. Astrophys.* **536**, A53 (2011)
4. Cuevas-Tello, J.C., Tiño, P., Raychaudhury, S.: How accurate are the time delay estimates in gravitational lensing? *Astron. Astrophys.* **454**, 695–706 (2006)
5. Cuevas-Tello, J.C., Tiño, P., Raychaudhury, S., Yao, X., Harva, M.: Uncovering delayed patterns in noisy and irregularly sampled time series: an astronomy application. *Pattern Recogn.* **43**(3), 1165–1179 (2009)
6. Fassnacht, C.D., Xanthopoulos, E., Koopmans, L.V.E., Rusin, D.: A determination of H_0 with the CLASS gravitational lens B1608+656 III. A significant improvement in the precision of the time delay measurements. *Astrophys. J.* **581**, 823–835 (2002)
7. Fathi-Vajargah, B., Khoshkar-Foshtomi, H.: Simulating nonhomogeneous poisson point process based on multi criteria intensity function and comparison with its simple form. *J. Math. Comput. Sci. (JMCS)* **9**(3), 133–138 (2014)
8. Greene, Z.S., Suyu, S.H., Treu, T., Hilbert, S., Auger, M.W., Collett, T.E., Marshall, P.J., Fassnacht, C.D., Blandford, R.D., Bradač, M., Koopmans, L.V.E.: Improving the precision of time-delay cosmography with observations of galaxies along the line of sight. *Astrophys. J.* **768**(1), 39 (2013)
9. Hainline, L.J., Morgan, C.W., Beach, J.N., Kochanek, C.S., Harris, H.C., Tilleman, T., Fadely, R., Falco, E.E., Le, T.X.: A new microlensing event in the doubly imaged Quasar Q 0957+561. *Astrophys. J.* **744**(2), 104 (2012)
10. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* **27**(2), 83–85 (2005)
11. Lewis, P.A., Shedler, G.S.: Simulation of nonhomogeneous poisson processes by thinning. *Nav. Res. Logistics Q.* **26**(3), 403–413 (1979)
12. Nawrot, M., Aertsen, A., Rotter, S.: Single-trial estimation of neuronal firing rates: from single-neuron spike trains to population activity. *J. Neurosci. Meth.* **94**(1), 81–92 (1999)
13. Park, B.U., Marron, J.S.: Comparison of data-driven bandwidth selectors. *J. Am. Stat. Assoc.* **85**(409), 66–72 (1990)
14. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**(3), 1065–1076 (1962)
15. Rasch, G.: The poisson process as a model for a diversity of behavioral phenomena. In: *International Congress of Psychology*, vol. 2, p. 2 (1963)
16. Refsdal, S.: On the possibility of determining Hubble’s parameter and the masses of galaxies from the gravitational lens effect. *MNRAS* **128**, 307 (1964)
17. Ross, S.M.: *Introduction to Probability Models*. Academic press, Boston (2014)
18. Rubinstein, R.Y., Kroese, D.P.: *Simulation and the Monte Carlo Method*, vol. 707. Wiley, New York (2011)
19. Shimazaki, H., Shinomoto, S.: Kernel bandwidth optimization in spike rate estimation. *J. Comput. Neurosci.* **29**(1–2), 171–182 (2010)
20. Sigman, K.: Poisson processes and compound (batch) poisson processes. *Lecture Notes*. Columbia University, USA (2007). <http://www.columbia.edu/ks20/4703-Sigman/4703-07-Notes-PP-NSPP.pdf>
21. Suyu, S.H., Auger, M.W., Hilbert, S., Marshall, P.J., Tewes, M., Treu, T., Fassnacht, C.D., Koopmans, L.V.E., Sluse, D., Blandford, R.D., Courbin, F., Meylan, G.: Two accurate time-delay distances from strong lensing: implications for cosmology. *Astrophys. J.* **766**(2), 70 (2013)