1   Background

- the time delay challenge 2014
- LSST

2   How do the data look like?

- data generation
- ground truth information
- FIG. one centered and one raw LC pair with and without error bars
- FIG. Histograms
- XX pair systems and XX quadratic ones - we ignore the quadratic systems (for this phase)

3   Our solution

1. smooth each observing window for each light curve individually (GPR using tau and sig from the ground truth – but possible to optimize the smoothing by cross validation as well. So we not only solve the problem of the raw time series being unevently-sampled in time, but also keep error estimate in this step)

2. measure cross correlation between the two light curves within each window and mark the timeshift that results in the maximum correlation coefficient as the estimated time delay of that window

3. compare the estimated time delays of different windows of the same pair - How though??

4. Problems with time delay estimation directly from correlation coefficients

   - each light curve has experienced microlensing events that are independent of those of the other image – solution: account for microlensing events in a statistical way rather than individually – the far-fetched proposed solution: train a classifier (maybe a neural network?) on the microlensing events and subtract a range of typical microlensing events from both signals, and use cross validation to choose the best microlensing effect by maximizing the cross correlation coefficients

   - we have no information about the effect of microlensing effects, they could only influence the trend of the data, so may be removed by detrending the light curves before cross correlating them, but the added peaks are impossible to be accounted for indicially, but only in a statistical sense – solution: work with centred, normalized, and detrended light curves

   - if the true time delay is close to the window size, the signal overlap with the correct timeshift is, so in our cases where the two signals are not identical (only time shifted), the correlation coefficient vs time shift won't show a significant peak – solution: account for this in dt errors

5. weight the estimated time delay at each window by the maximum correlation coefficient

6. use a clustering algorithm (nearest neighbour regression maybe?) on estimated dts from different windows, to derive out estimation for the pair

7. use a regression algorithm with the maximum correlation coefficients and MSE of hte two signals as features for the entire pair (where true dt, as well as other ground truth information we have are the labels and we hope to do a multi-dimensional regression!)

8. propagate errors – no solution (yet!!)

4   How far along are we?

mentioned above as well!

5   What is next?

References