OUTBURSTS WITH EXPONENTIAL MOVING AVERAGES

Roy D. Williams
Royal Observatory, Edinburgh EH9 3HJ, Scotland
roy@roe.ac.uk

ABSTRACT

As we near first light of the LSST[1], astronomers are keen to exploit the deluge of observations of transient sources. In this paper we consider how to detect *outburst* rapidly and reliably, using the Exponential Moving Average (EMA). In the Lasair-ZTF project[2], EMAs of both g and r filter are computed and stored for 2-day, 8-day, and 28-day timescales. We give mathermatical theory, present some experiments, and show outbursts from the ZTF survey[3] detected using the EMA criteria.

1 Introduction

The ZTF survey[3] is providing a data-firehose of many millions of detections of transient and variable astronomical objects, and in a few years the LSST survey[1] will increase the rate of the firehose by a factor of 50 or more. The challenge for astronomers is to extract useful science from the data, and a special challenge is to do so in real time. When an astrophysical object does something interesting, it is often true that the earlier that interest is detected, the quicker the follow-up observations can be initiated, and the better the quality of the resulting science.

There are many ways to compute 'variability indices' from a light curve, and many of these find features that have detections both before and after the feature. However, we are also interested in *early classification*, where emergent features are detected before their full elaboration: examples are the outburst of a source such as a protoplanteary disk, a binary with an accretion disk, or an active galactic nucleus. There are such measures, based on deep learning and machine learning, however our fucus here is to provide a simple and intuitive way to determine if aa source magnitude is rising.

First we describe the Exponential Moving Average (EMA) and its mathematical properties, then describe some experiments that detect a step or a ramp increas in the presence of noise. Finally we show some outbursts of cataclysmic variables, discovered in ZTF data using EMAs. We can also note that the EMA is a primary tool in financial markets[5][6], where traders are interested in knowing whn a commodity price is increasing or decreasing in oreder to buy or sell and make a profit.

2 Exponential Moving Average

Suppose now is time T, we have a signal s_i that has been measured at irregular intervals in the past:

$$t_0 < t_1 < ... < t_n = T \tag{1}$$

where the most recent measurement s_n has been at time T. Each signal measurement has an error estimate σ, but we shall assume the error to be the same for each measurement. For a timescale τ, we can define the EMA S_n recursively. First we define:

$$E_i \equiv e^{-(t_n - t_{n-i})/\tau} \tag{2}$$

then the EMA can be defined recursively as:

$$S_n = S_{n-1}E_1 + s_n(1 - E_1) \tag{3}$$

From this, we can write explicitly the following two expressions:

$$S_n = \sum_{i=0} s_{n-i}(E_i - E_{i+1}) \tag{4}$$

$$S_n = s_n - \sum_{i=1} E_i(s_{n-i+1} - s_{n-1}) \tag{5}$$

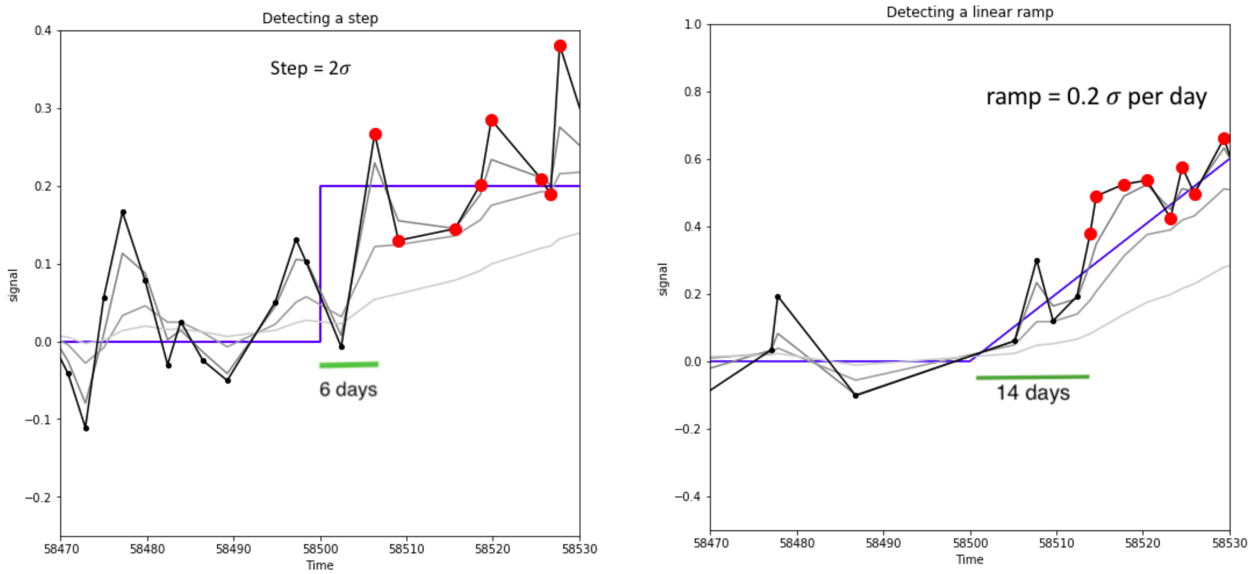

Figure 1: Detection of (left) a step and (right) a linear ramp. The signal (black with markers) is a sum of the ramp (blue) and gaussian noise, and is sampled irregularly. Three Exponential Moving Averages are shown with timescales of 2, 8, and 28 days in grey. The red markers show where outburst is detected by differences of the EMA values; the criterion for a point depends only on the signal and three EMA values at that point. With the error of the signal at 0.1, the step is size 2σ and the ramp has rate 0.2σ per day. The step is detected in 6 days, the ramp in 14 days.

2.1 Continuous Limit

Consider the continuous limit, where s_i are samples from a differentiable function $s(t)$ that changes slowly between sample times. Then equation (5) becomes an integral:

$$S(T) = s(T) - \int_{-\infty}^{T} \frac{ds}{dt} e^{-(T-t)/\tau} dt \tag{6}$$

Integrating by parts and changing t to $T - t$, this becomes a Laplace transform of the past history $t < T$ of the function $s(t)$:

$$S(T) = \frac{1}{\tau} \int_{0}^{\infty} s(T-t) e^{-t/\tau} dt \tag{7}$$

The Laplace transform[8] is *invertible*, meaning that the full history of the original signal can be reproduced from all the EMAs at time T, for every timescale.

Let us compute the EMAs for some simple signals. First the step function that occurs at time $T - q$, or q before time T, where $s(t) = 1$ iff $T - q < t < T$, so the EMA is:

$$S(T) = \frac{1}{\tau} \int_{0}^{q} e^{-t/\tau} dt = (1 - e^{-q/\tau}) \tag{8}$$

The ramp function has a linear increase starting at time $T - q$, where $s(t) = (t - (T - q)$ iff $T - q < t < T$, so the EMA is:

$$S(T) = \frac{1}{\tau} \int_{0}^{q} (q-t) e^{-t/\tau} dt = q - \tau(1 - e^{-q/\tau}) \tag{9}$$

2.2 EMA for random variables

If the signal s_i is a measurement, it may be convenient to think of an ensemble of such measurements spread by an error estimate σ_i, in other words a random variable. The EMAs will then also be random variables, and we can use equations (3-5) for their means. We can get the variance of the EMAs from (3) recursively:

$$\mathrm{Var} S_n = {E_1}^2 \mathrm{Var} S_{n-1} + {\sigma_i}^2 (1 - E_1)^2 \tag{10}$$

and from (4)

$$\mathrm{Var} S_n = \sum_{i=0} {\sigma_{n-i}}^2 (E_i - E_{i+1})^2 \tag{11}$$

In the case where each measurement has the same error estimate, then equation (11) reduces to:

$$\mathrm{Var} S_n = \sigma^2 \sum_{i=0} (E_i - E_{i+1})^2 \tag{12}$$

which depends only on the spacing of the sample times. In the case of uniform sample time Δ, we have

$$\mathrm{Var} S_n = \sigma^2 \sum_{i=0} (E_i - E_{i+1})^2 = \sigma^2 \frac{1 - E_1}{1 + E_1} = \sigma^2 \frac{1 - e^{-\Delta/\tau}}{1 + e^{-\Delta/\tau}} \tag{13}$$

Thus in the limit where the timestep is much less than the timescale, the variance of the EMA is $\Delta/2\tau$ times the variance of the signal.

3 Experiments

In the following we describe some experiments with using EMAs to determine if a signal is rising/outbursting. We consider a signal which is a sum of a Gaussian distribution and either a step or a linear ramp. We attempt to detect the step/ramp at each sample, using the signal itself, and three EMAs to represent the past behaviour of the signal.

The intervals between observations are distributed as an exponential distribution, but excluding interval sizes near zero; the intervals are 0.5 + an exponential distribution of mean 2.5, so that the mean intervasl between observations is 3, which is the cadence of the ZTF survey[4]. We choose EMAs with time scales 2, 8, and 28 to detect the step or ramp. We used a combination of these to decide when the signal is rising, and not just a fluctuation:

```
E02 > E08 > E28 + L
```

where L is a multiple of the signal error $sigma$. Results are shown in Figure 2, that plots two things we would want minimised: the time to detect the outburst (vertical axis), and the rate of false positives where the signal has no imposed outburst. Each curve is for a specific size of step or ramp, and each curve has L varying along it. The value $L = 1.0\sigma$ generally gives a false alarm rate below 0.01, together with detecting the signal in the first two observations.

It might be argued that a mich simpler detector of outburst would be to simply check if the difference from the mean is greater than 3σ. However, there are a few problems with this. First the mean signal value is not intrinsically known, and there may be long term variations; this means that a EMA with a long timescale would be a practical stand-in for that mean. It may be that the signal has non-gaussian statistics, so that the 3σ criterion is not justified. To be scientifically significant, we are interested here in multiple measurements that show outburst behaviour, a transition from low to high state, whereas a single outlier could be for technical causes in the measurment itself.

4 Outbursts from the ZTF survey

We have tried using the EMA criterion described above to look for outbursts in sources that have previously outburst. We took the 1038 cataclysmic variables that have previuos outbursts, from the catalogue of Coppejans et. al[9], and croessmatched them with transients from the ZTF[3], using the watchlist tool from Lasair[2]. Lightcurves are returned in two filters, g and r, and we computed with 2-day timescale, with 8-day timescale, and with 28-day timescale. We ran the EMA criterion with the magnitudes from each, using the attribute names that are used in the Lasair[2] broker. For finding detections (`candidate`) from the historical archive:

```
select objectId from candidates where
-dc_mag_g02 > -dc_mag_g08 and -dc_mag_g08 > -dc_mag_g28 + 0.3
and
-dc_mag_g02 > -dc_mag_g08 and -dc_mag_g08 > -dc_mag_g28 + 0.3
```

The latest values of the EMAs is copied into the `object` Note the minus signs before each quantity, this is because astronomical magnitude is an inverted scale. The latest values of the EMA are also available as attributes of the `object`, with a stream generated by this criterion:

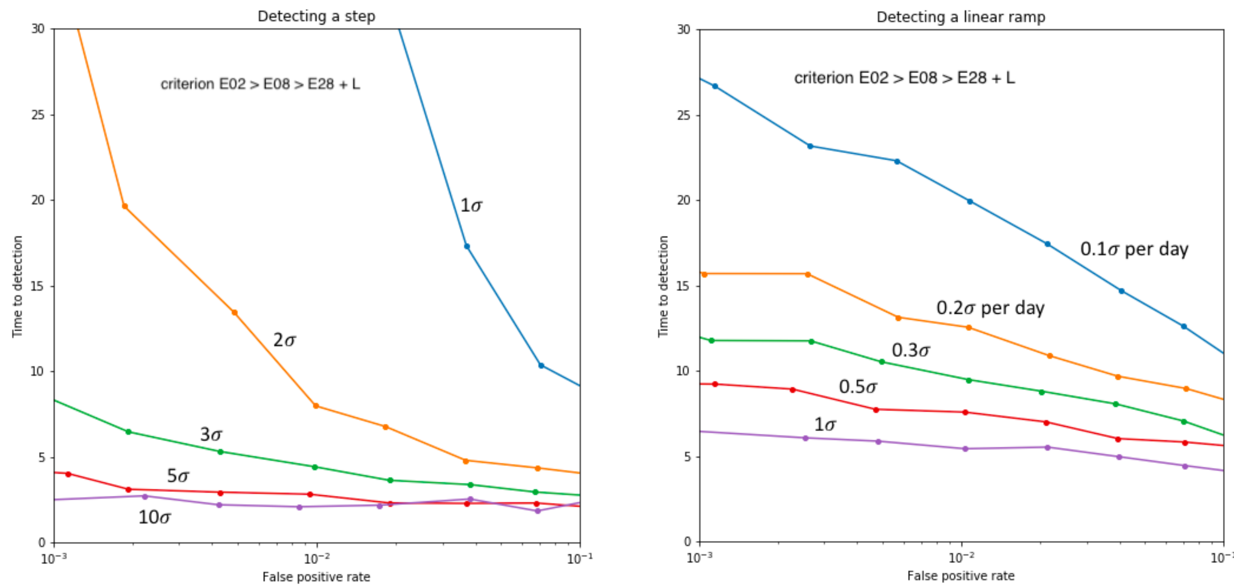

Figure 2: Performance of a detection criterion for the step outburst (left) and the ramp outburst (right). In each case, we are trying to minimise false positives, when random noise looks like an outburst, and also minimise the time to detection whne a real outburst has started. Each point of the curve is an average of 100 runs, varying the parameter L from 0.3 to 1.5 along the curve.

```
select objectId from objects where
-latest_dc_mag_g02 > -latest_dc_mag_g08 and -latest_dc_mag_g08 > -latest_dc_mag_g28 + 0.3
and
-latest_dc_mag_g02 > -latest_dc_mag_g08 and -latest_dc_mag_g08 > -latest_dc_mag_g28 + 0.3
```

For more information on the attributes available for queries in Lasair, see[7].

Figure 3 shows four examples outbursts, which can be seen in both filters. The number 0.3 is chosen to yield outbursts, yet keep the false positive rate low.

We should emphasise that the queries shown above are just an approximate guide to detecting outbutsts using the EMAs supplied with Lasair. Careful scientists should use the historical archive to build a query, before converting it to a real-time stream.

References

[1] LSST Alerts: key numbers `https://dmtn-102.lsst.io/`

[2] Lasair Transient Broker for the ZTF survey `https://https://lasair.roe.ac.uk/`

[3] Zwicky Transient Facility `https://www.ztf.caltech.edu/`

[4] Zwicky Transient Facility data release, Figure 3b in `https://www.ztf.caltech.edu/page/dr1`

[5] Lawrence Pines *How To Analyze Trends With Moving Average Exponential Ribbons* `https://commodity.com/technical-analysis/exponential-ribbons/`

[6] Dayo, *Predicting an Unpredictable Stock Market* `https://simplystocksandoptions.com/predicting-an-unpredictable-stock-market/`

[7] Schema for candidates, objects, and classifications in Lasair `https://lasair.roe.ac.uk/schema`.

[8] `https://en.wikipedia.org/wiki/Laplace_transform`

[9] D. L. Coppejans et. al, *Statistical properties of dwarf novae-type cataclysmic variables: the outburst catalogue*, MNRAS 456 (2016) p.4441-4454. Available in Vizier J/MNRAS/456/4441.

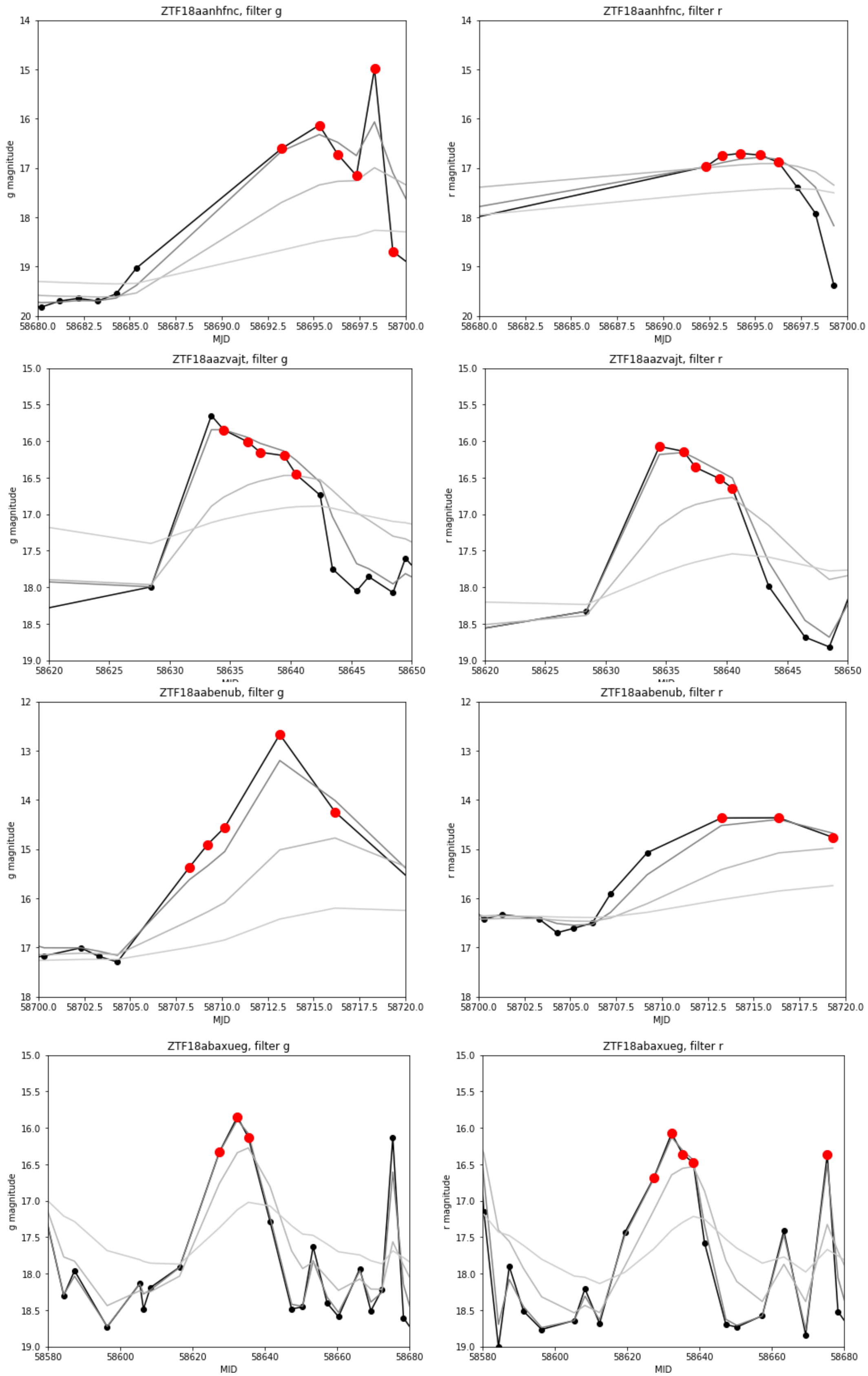

Figure 3: Detections of outburst from the ZTF survey, as reported from the Lasair transient broker.