

Anomaly Detection with Double Exponential Smoothing

Preetam Jinka

May 31, 2015

1 Introduction

1.1 Motivation

Anomaly detection in time series is used to detect observations that significantly deviate from a usual pattern or range. This is extremely important for monitoring systems. A major issue in monitoring systems is that they deal with many time series metrics with varying and complicated characteristics. It is unreasonable to fit accurate models to every time series to detect anomalous observations. Anomaly detection in time series is also done online as new observations of a process arrive. As a result, recursive algorithms are necessary. We propose double exponential smoothing as a method of addressing this problem. Double exponential smoothing provides a method for adaptive anomaly detection of time series with minimal assumptions made on the input data.

1.2 Holt-Winters

The Holt-Winters method is generally known as a method of triple exponential smoothing. The exponential moving average filter applied three times accounts for level, trend, and seasonality for a given time series. The single exponentially weighted moving average is defined as

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1}. \quad (1)$$

We can introduce another smoothing parameter, β , to account for a linear trend (citation??).

$$\begin{aligned} s_t &= \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \end{aligned} \quad (2)$$

Finally, to account for a multiplicative seasonality, we introduce another smoothing parameter, γ .

$$\begin{aligned} s_t &= \alpha \frac{x_t}{g_{t-L}} + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \\ g_t &= \gamma \frac{x_t}{s_t} + (1 - \gamma)g_{t-L} \end{aligned}$$

For our purposes, we will not consider seasonality.

1.3 “Anomaly detection”

We define “anomalies” in time series as statistically unlikely values given a limited number of previous terms. The process of detecting these anomalies occurs every time a new value is observed.

2 One step-ahead prediction interval

In the context of anomaly detection, we are interested in detecting points that fall outside previously computed prediction intervals. For simplicity, we consider an AR(1) time series and compared the AR prediction interval with the Holt-Winters approach.

2.1 Simulations

The AR(1) process is given by

$$x_t = \phi x_{t-1} + \varepsilon_t.$$

We first generate N terms of an AR(1) process using the `arima.sim` function in R. The smoothed time series, s_t , is then computed iteratively using either (1) or (2).

The square-error time series is defined as

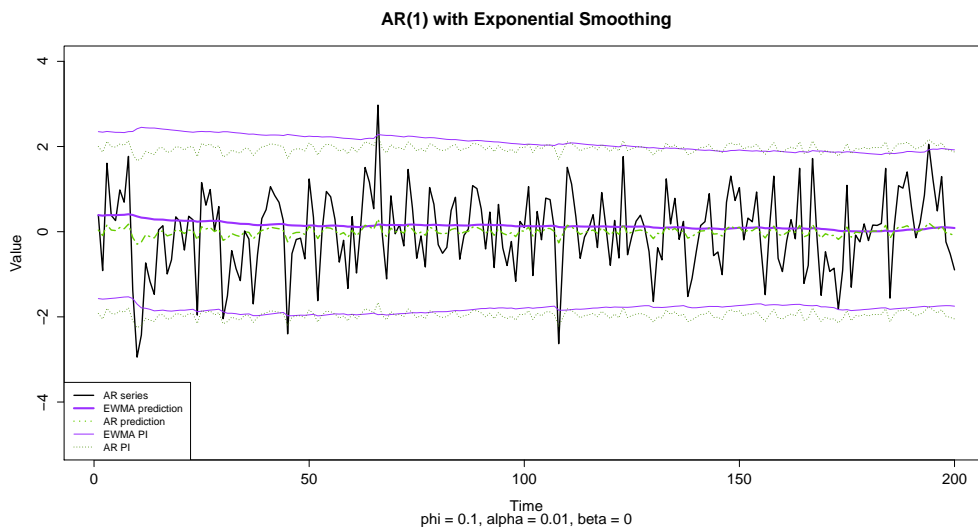
$$e_t = (1 - \alpha)e_{t-1} + \alpha(s_t - x_t)^2.$$

e_t serves as an exponentially-weighted moving MSE.

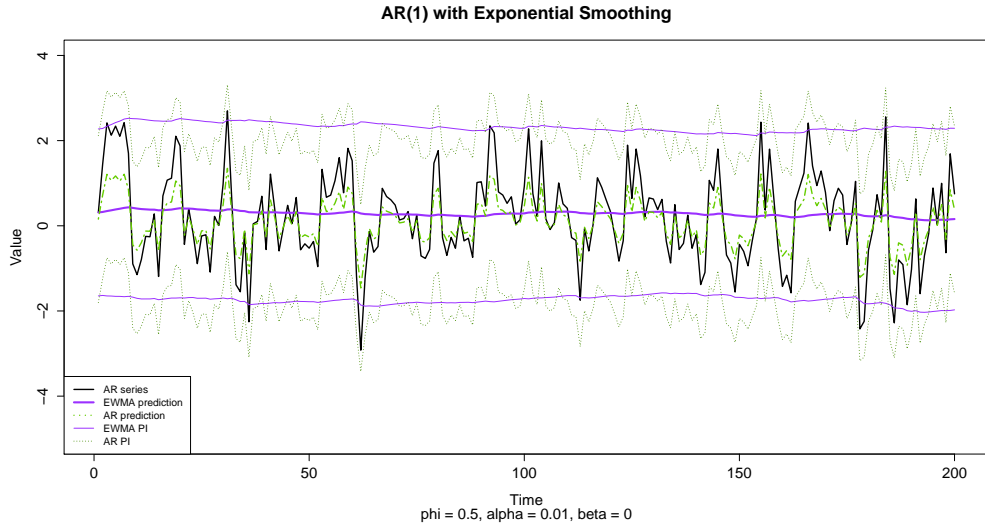
The prediction intervals are computing using

$$s_t \pm 1.96\sqrt{e_t}. \quad (3)$$

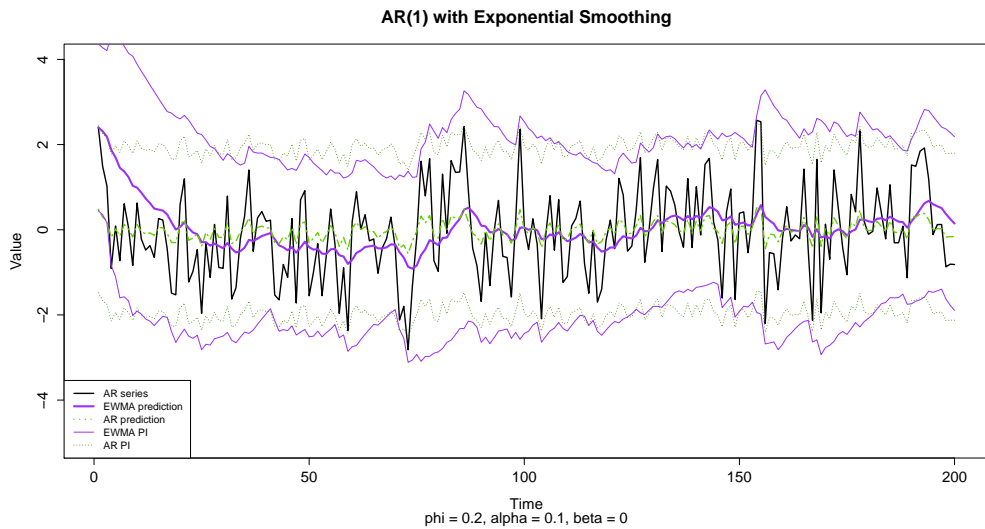
The following are plots of a simulated AR(1) with single and double exponential smoothing. Each plot also contains the AR(1) and exponentially-weighted moving average (EWMA) predicted values, and the respective prediction intervals.



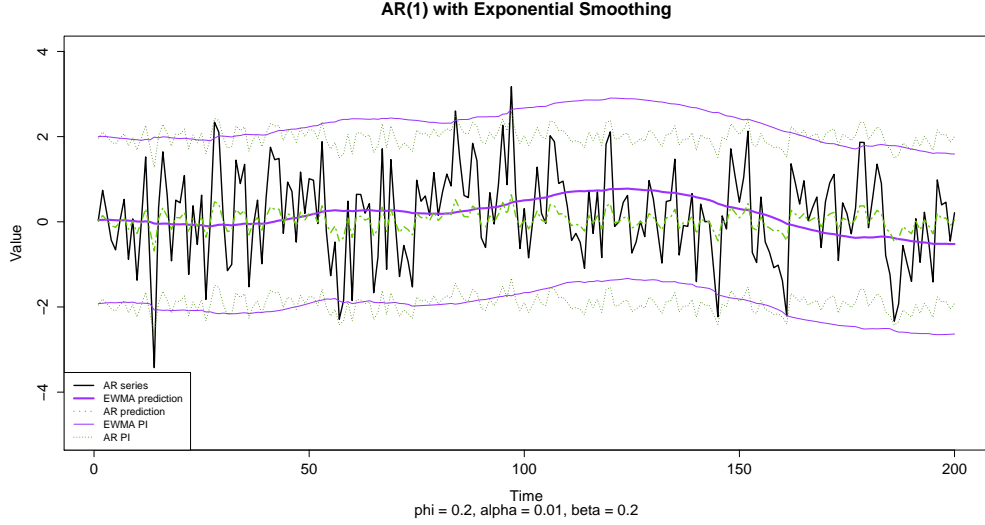
The first plot only has single exponential smoothing, i.e. the value of β is zero. The first feature that should be noted is that the predicted value from the smoothed time series, labeled “EWMA prediction”, eventually approaches the predicted value for the AR process. The second significant feature is that the EWMA prediction interval also approaches the interval from the one step-ahead AR forecast.



In the second plot, the EWMA weights α and β are unchanged from the first plot but the ϕ coefficient of the simulated AR series is increased to 0.5. Observe that both the confidence interval lines and predicted values of the smoothed series deviate significantly from the values calculated using the corresponding AR model.



In the third plot, the coefficient ϕ is reduced to 0.2 and the smoothing parameter α is increased to 0.1, which increases the influence of newer values. Observe that in this case, the smoothed values do not deviate as significantly as in the previous example, but seem to have higher variability.



In the final plot, we decrease α to 0.01 but introduce the β parameter set to 0.2. The non-zero β parameter introduces a non-zero trend in the smoothed series, which is why the smoothed values and their corresponding prediction intervals start to move away from a zero mean at around timestamp 100.

2.2 Coverage probabilities

The coverage probabilities were tested against the AR(1) process described in the previous section. In order to calculate the coverage probability, we calculate the number of times a newly observed value falls inside of the predicted interval for that value. First, we simulate an AR(1) process as before. We run the simulation and stop at least one step before the end. Finally, we compare the next simulated value against the prediction interval of the smoothed series using (3). We increment a counter if the value falls inside of the prediction interval. This process is repeated for at least 5000 iterations.

As the table shows, the choices of α and β greatly affect the overall coverage probabilities. In practice, many combinations of these parameters should be checked in order to select optimal values.

| Iterations | α | β | Expected | Actual | Absolute difference |
|------------|----------|---------|----------|--------|---------------------|
| 5000 | 0 | 0 | 0.05 | 0.165 | 0.115 |
| 5000 | 0.05 | 0 | 0.05 | 0.0568 | 0.0068 |
| 5000 | 0.05 | 0.05 | 0.05 | 0.0528 | 0.0028 |
| 10000 | 0.25 | 0.05 | 0.05 | 0.0876 | 0.0376 |
| 10000 | 0.25 | 0.25 | 0.05 | 0.0857 | 0.0357 |

3 Smoothed MA(1)

Let x_t be MA(1), which is defined as

$$x_t = w_t + \theta w_{t-1}.$$

$$\begin{aligned}
\mathbb{E}(x_t) &= 0 \\
\text{Var}(x_t) &= (\theta^2 + 1)\sigma^2 \\
\text{Cov}(x_t, x_{t+h}) &= \begin{cases} \sigma^2(1 + \theta^2), & \text{if } h = 0. \\ \sigma^2\theta, & \text{if } h = 1. \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{4}$$

We define s_t as x_t with single exponential smoothing with parameter α , i.e.

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1}. \tag{5}$$

3.1 Iterative form of s_t

Using (14), we calculate the first few terms of s_t .

$$\begin{aligned}
s_1 &= (1 - \alpha)s_0 + \alpha x_1 \\
s_2 &= [(1 - \alpha)s_0 + \alpha x_1](1 - \alpha) + \alpha x_2 \\
&= (1 - \alpha)^2 s_0 + \alpha(1 - \alpha)x_1 + \alpha x_2 \\
s_3 &= [(1 - \alpha)^2 s_0 + \alpha(1 - \alpha)x_1 + \alpha x_2](1 - \alpha) + \alpha x_3 \\
&= (1 - \alpha)^3 s_0 + \alpha(1 - \alpha)^2 x_1 + \alpha(1 - \alpha)x_2 + \alpha x_3
\end{aligned}$$

This can be rewritten in iterative form as

$$s_t = (1 - \alpha)^t s_0 + \alpha \sum_{i=1}^t (1 - \alpha)^{t-i} x_i.$$

3.2 Variance of s_t

We will derive the closed-form variance of s_t . Note that s_0 is assumed to be constant in the following equation.

$$\begin{aligned}
\text{Var}(s_t) &= \text{Var}\left((1 - \alpha)^t s_0 + \alpha \sum_{i=1}^t (1 - \alpha)^{t-i} x_i\right) \\
&= \text{Var}\left(\alpha \sum_{i=1}^t (1 - \alpha)^{t-i} x_i\right)
\end{aligned} \tag{6}$$

Recall the following identity

$$\text{Var}\left(\sum_{i=1}^t x_i\right) = \sum_{i=1}^t \text{Var}(x_i) + \sum_{i \neq j} \text{Cov}(x_i, x_j) \tag{7}$$

Using (7), we can rewrite (6) as

$$\begin{aligned}
\text{Var}(s_t) &= \text{Var}\left(\alpha \sum_{i=1}^t (1 - \alpha)^{t-i} x_i\right) \\
&= \sum_{i=1}^t \text{Var}(\alpha(1 - \alpha)^{t-i} x_i) + \sum_{i \neq j} \text{Cov}[\alpha(1 - \alpha)^{t-i} x_i, \alpha(1 - \alpha)^{t-j} x_j]
\end{aligned} \tag{8}$$

We compute the two terms separately starting with the left.

$$\sum_{i=1}^t \text{Var}(\alpha(1-\alpha)^{t-i}x_i) = \sum_{i=1}^t \alpha^2(1-\alpha)^{2t-2i} \text{Var}(x_i)$$

We can substitute the variance of x_i using (4).

$$= \sigma^2(\theta^2 + 1) \sum_{i=1}^t \alpha^2(1-\alpha)^{2t-2i} \quad (9)$$

For the right term, we arrive at the following

$$\sum_{i \neq j} \text{Cov}[\alpha(1-\alpha)^{t-i}x_i, \alpha(1-\alpha)^{t-j}x_j] = \sum_{i \neq j} \alpha^2(1-\alpha)^{2t-i-j} \text{Cov}(x_i, x_j) \quad (10)$$

Observe that using (4), we see that the only terms that contribute are those with i and j differing by 1. Due to symmetry, each of those terms contributes twice. Therefore, (10) can be rewritten as

$$\sum_{i \neq j} \text{Cov}[\alpha(1-\alpha)^{t-i}x_i, \alpha(1-\alpha)^{t-j}x_j] = 2\sigma^2\theta \sum_{i=1}^{t-1} \alpha^2(1-\alpha)^{2t-(2i+1)}. \quad (11)$$

Combining (9) and (11), we get

$$\begin{aligned} \text{Var}(s_t) &= \text{Var}\left(\alpha \sum_{i=1}^t (1-\alpha)^{t-i}x_i\right) \\ &= \sigma^2(\theta^2 + 1) \sum_{i=1}^t \alpha^2(1-\alpha)^{2t-2i} + 2\sigma^2\theta \sum_{i=1}^{t-1} \alpha^2(1-\alpha)^{2t-(2i+1)} \end{aligned} \quad (12)$$

□

3.3 Empirical results

The previously discussed simulation was modified to use a simulated MA(1) using `arma.sim`. For n iterations, the t -th value of the smoothed time series s_t was collected. The variance of these collected values was compared to the closed-form variance (12) derived in the previous section.

| n | θ | σ | t | α | Empirical variance | Closed-form variance | Absolute difference |
|-------|----------|----------|-----|----------|--------------------|----------------------|---------------------|
| 20000 | 0.1 | 1 | 1 | 1 | 1.006463 | 1.010000 | 0.00354 |
| 20000 | 0.1 | 1 | 1 | 0.5 | 0.2543537 | 0.2525000 | 0.00185 |
| 20000 | 0.1 | 1 | 10 | 0.5 | 0.3648053 | 0.3699996 | 0.00519 |
| 20000 | 0.1 | 1 | 10 | 0.1 | 0.05575114 | 0.05474687 | 0.001 |
| 20000 | 0.1 | 1 | 20 | 0.5 | 0.3649438 | 0.3700000 | 0.00506 |
| 20000 | 0.1 | 1 | 50 | 0.5 | 0.3778751 | 0.3700000 | 0.00788 |

4 Smoothed MA(2)

Let x_t be MA(2), which is defined as

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}.$$

$$\begin{aligned}
E(x_t) &= 0 \\
\text{Var}(x_t) &= \sigma^2(1 + \theta_1^2 + \theta_2^2) \\
\text{Cov}(x_t, x_{t+h}) &= \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2), & \text{if } h = 0. \\ \sigma^2(\theta_1 + \theta_1\theta_2), & \text{if } h = 1. \\ \sigma^2\theta_2, & \text{if } h = 2. \\ 0, & \text{otherwise.} \end{cases} \quad (13)
\end{aligned}$$

As with the MA(1) case, we define s_t as x_t with single exponential smoothing with parameter α , i.e.

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1}. \quad (14)$$

Recall from (8) that we can express $\text{Var}(s_t)$ as

$$\text{Var}\left(\alpha \sum_{i=1}^t (1 - \alpha)^{t-i} x_i\right) = \sum_{i=1}^t \text{Var}(\alpha(1 - \alpha)^{t-i} x_i) + \sum_{i \neq j} \text{Cov}[\alpha(1 - \alpha)^{t-i} x_i, \alpha(1 - \alpha)^{t-j} x_j].$$

We repeat the procedure applied for the MA(1) case. The first term on the right hand side becomes

$$\sum_{i=1}^t \alpha^2 (1 - \alpha)^{2t-2i} \text{Var}(x_i) = \sum_{i=1}^t \alpha^2 (1 - \alpha)^{2t-2i} (\sigma^2(1 + \theta_1^2 + \theta_2^2)) = \sigma^2(1 + \theta_1^2 + \theta_2^2) \sum_{i=1}^t \alpha^2 (1 - \alpha)^{2t-2i}.$$

For the second term, we have

$$\sum_{i \neq j} \alpha^2 (1 - \alpha)^{2t-i-j} \text{Cov}(x_i, x_j) = 2\sigma^2(\theta_1 + \theta_1\theta_2) \sum_{i=1}^{t-1} \alpha^2 (1 - \alpha)^{2t-(2i+1)} + 2\sigma^2\theta_2 \sum_{i=1}^{t-2} \alpha^2 (1 - \alpha)^{2t-(2i+2)}.$$

Therefore,

$$\begin{aligned}
\text{Var}(s_t) &= \sigma^2(1 + \theta_1^2 + \theta_2^2) \sum_{i=1}^t \alpha^2 (1 - \alpha)^{2t-2i} \\
&\quad + 2\sigma^2(\theta_1 + \theta_1\theta_2) \sum_{i=1}^{t-1} \alpha^2 (1 - \alpha)^{2t-(2i+1)} \\
&\quad + 2\sigma^2\theta_2 \sum_{i=1}^{t-2} \alpha^2 (1 - \alpha)^{2t-(2i+2)}.
\end{aligned}$$

5 Smoothed MA(q)

$$\begin{aligned}
\text{Var}(s_t) &= \text{Var}(x_t) \sum_{i=1}^t \alpha^2 (1 - \alpha)^{2t-2i} \\
&\quad + 2 \sum_{h=1}^{\infty} \text{Cov}(x_t, x_{t-h}) \alpha^2 (1 - \alpha)^{2t-(2i+h)}
\end{aligned}$$

References

- [1] Prajakta S. Kalekar, *Time series Forecasting using Holt-Winters Exponential Smoothing*. Kanwal Rekhi School of Information Technology, 2004.