**Simple variance estimator for data stream mining**

**July 2012**

**Hunor Albert-Lorincz**

**Abstract.** *This paper introduces a rolling variance estimator that can be computed with a single pass over the dataset, considering every data point only once. It is computationally efficient and has good properties.*

Consider a continuous infinite data stream with elements *Y1, Y2, ..*

The standard exponential moving average with the shape parameter is defined by the recursion , .

Smaller values of the parameter α give smoother averages, while higher values are more reactive and follow the data points closer. This behaviour is illustrated on figure 1.

***Figure 1****. Comparison of standard rolling average (last 30 points) and exponential moving averages (EMA) with different shape parameters. Higher values of a are more reactive to changes in data.*

We introduce the **exponential moving variance** defined by the recursion

With shape parameters α, β, and initial value .

Empirically, the exponential moving variance follows very closely the standard rolling variance, as illustrated in figure 2.

***Figure 2.*** *The exponential moving variance (α=0.09, β=0.1) follows closely the rolling variance (last 30 points used).*

*TODO*

* Is EMV a non biased estimator for the variance ???
* Compare distribution of errors between *EMVi* and *Yi* to *σi* and *Yi*.

*The covariance of X and Y is given by*

*We attempt to establish a recursion*

**References**