The Exponential Moving Average

© Malcolm Moore
14-Dec-2003

(2003, 2007, 2009, Sept 2012)

Exponential Charge and Decay Explained

In "Physics land", all natural actions follow an exponential rate of charge and / or decay, where a change initially happens fast and as time goes on the change slows down and finally appears to settle into a new position or state of being.

One typical example of an exponential decay is when a tank is full of water and the tap at the bottom is opened the water initially rushes out, due to the height of the water above the tap. As the height of the water over the tap falls, then so does the pressure and the rushing flow settles down to a flow that is considerably weaker. As the height of the water over the tap continues to fall, the rate of water flow through the tap gets weaker and weaker until it is but a drip.

Another typical example is when car headlights are turned off, they take about a second to dim out and go dark as the filament cools down to room temperature in an exponential manner.

Exponential charges and decays are extremely common and it is a natural phenomenon everywhere! Another classical example is when rain starts and stops falling, rain density over time is an exponential function, and it follows the same exponential decay rules!
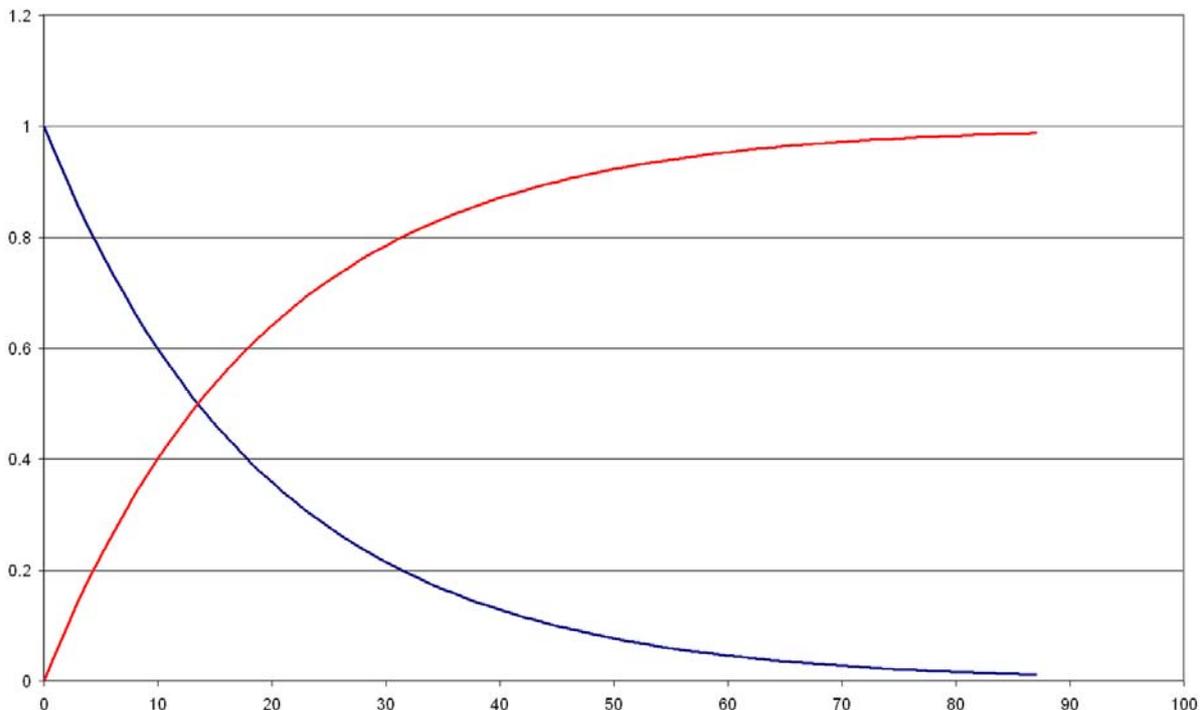
The graph / chart above shows a typical exponential decay in deep blue starting at 1 on the left hand side as time increases from left to right. The value exponentially falls away towards zero and theoretically never reaches zero until infinity time. The red exponential charge curve on the same graph starts at zero at time zero, and with an exponential curve asymptotically reaches the value of 1 at time infinity.

In "Electronics Land", analogue values are time related by the initial charge, the exponential charge or discharge rate, and comparative time. These values follow in a family of curves like the ones above for a charge or discharge.

For a discharge (starting from 1), the value of an exponential to a negative power can be described by the function ($\exp^{-t}$), where t =1, 2, 3, 4, 5, etc., and it follows that $\exp^0 = 1$, $\exp^{-1} = 0.367879$, $\exp^{-2} = 0.1353352$ etc.

For a charge (starting from 0 and asymptotically reaching 1), the value of an exponential to a negative power can be described by the function ($1 - \exp^{-t}$), where t =1, 2, 3, 4, 5, etc., and it follows that ($1 - \exp^0$) = 0, ($1 - \exp^{-1}$) = 0.6321205, ($1 - \exp^{-2}$) = 0.8646647 etc.

In practice, to relate the exponential charge or discharge instantaneous value with time, it is necessary to scale the time factor, and that is done with what is called a **Time Constant**, which is positioned as a ratio of the power of the exponential "e or exp" with the scaling "T", called "tor" in Greek.

So the Charge curve gets the following scaling ($1 - \exp^{-t/T}$) and the Discharge curve has the following scaling ($\exp^{-t/T}$). With analogue electronics the component values are deliberately scaled such that they exactly match the associated **Time Constant**.

So now, with the **Time Constant** concept any simple exponential time related charge or discharge value can be calculated and accurately predicted.

The formula for the above discharge chart is simple: Amplitude = exp ^(-t/19.876) where t is the value of the X axis and T = 19.876. For the charge curve, the value of T is identical but the formula is Amplitude = 1 - exp ^(-t/19.876).

It is this characteristic of an exponential "lazy response" that can be mathematically engineered in to follow a time changing sequence of events.

Electronic Filter Theory is a very closely related Engineering topic that studies the configurations of circuits that have Time Constants. Electronic filters are usually described in the frequency domain, not the time domain, which is extensively used in Technical Analysis.

In this case, where there is only one Time Constant, this is called an analogue First Order Low Pass Filter.

We now can recognise that the trade price noise caused by emotional traders consistently varying the trade prices, appears as higher frequencies in comparison to the general trend. A Low Pass filter that is engineered to have its Stop Band in these higher frequencies will severely attenuate this trading noise, leaving a general trend that makes decisions with considerably ease than without the filter.

Up until this point we have exclusively used analogue exponentials to outline analogue filter properties. The problem is that the securities market is a time sequential series of trades at varying prices for particular securities, and using an analogue filter technique to reduce the noise (trade price variation) is highly impracticable because we have the data saved in computers and we would like the results to be displayed statically.

Computers are really good at number crunching, holding and displaying data in a number of forms including graphically.

With all these above considerations it makes very practical sense to use a digital filter, but before we get into that we need to understand how digital exponentials are generated as they are one of the essential components in Exponential Moving Average filters used in Technical Analysis.

Digital Exponentials Explained

The analogue exponential is a continuous curve, and the amplitude can be calculated for any value of T. If the constant T is changed, then the decay rate will change accordingly. If T is made smaller, the decay rate is reduced.

With Technical Analysis, the times are clocked, and at those clocked times (Open, Close etc,) the price and other values are sampled. If we wish to use an exponential decay as a filtering tool then we need to adapt the exponential decay into a digital form and fortunately this is rather easy. See the chart below:
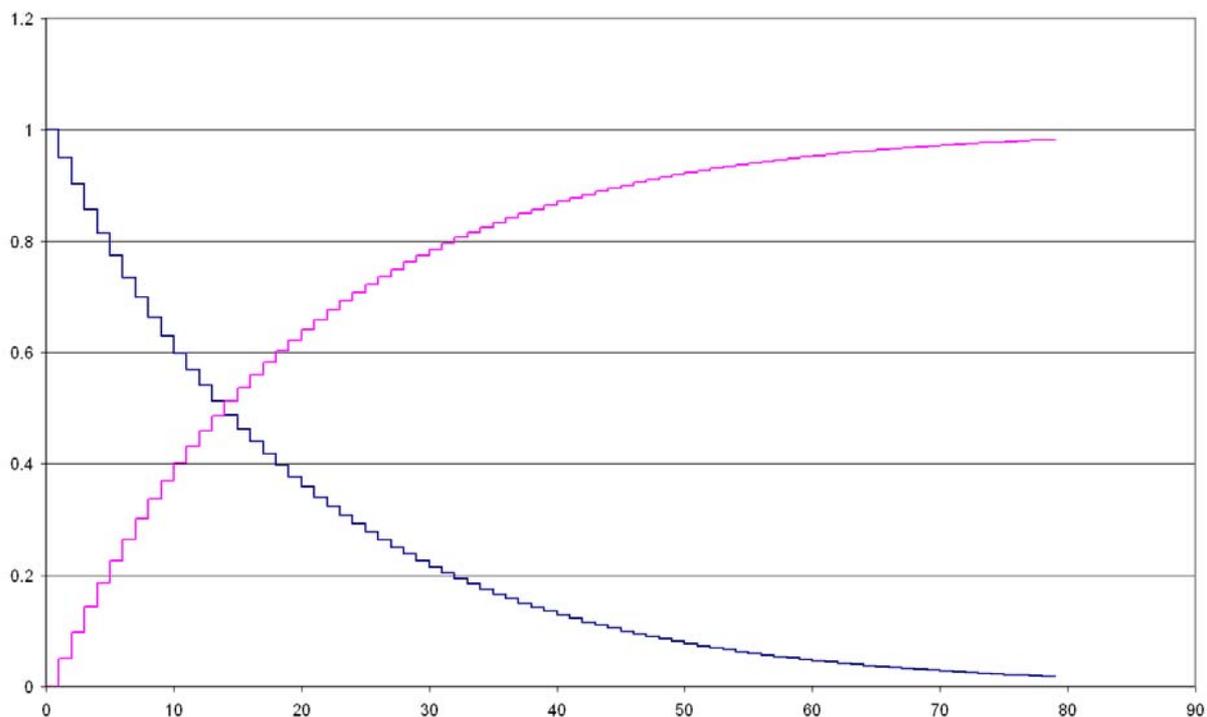
The formula for the decay curve in the above chart was created by having each next value as being 0.95 * the previous (recursive) value. So the amount of decay is really based on a fraction that is slightly less than one. Putting this in another view, the final value for each time period is a recursive value, that is, the new value is directly created (recurring) from the old value.

The formula for the charge curve is simply 1 minus the current decay curve value.

The digital (time clocked) equivalent of a continuous decay curve as shown in the first chart is a stepped response that "looks like" an exponential decay as shown in the first chart above. Yes, these two exponential shapes can be directly related, and yes, the relation process is relatively easy.

If we take the equation for the Charge curve in analogue terms we get

---

$Y(t) = 1 - \exp^{(-t/T)}$

Where Y(t) is the instantaneous value of the charge at time t. Now, the Count of Clocking Periods as the time factor "t" in the directly above equation, then we can arbitrarily call up any number of Count and work it through the equation.

Let us choose t = 39 (clocked) periods, so:

$y(39) = (1-\exp^{-39/T}) = 0.8646647$ so $0.1353352 = \exp^{-39/T}$ and

So, $\ln(0.1353352) = -2$ so $\exp^{-2} = \exp^{-39/T}$ so

$-2 = -39/T$, and transposing, T = 19.5

So what did all that High School maths mean?

It basically meant that the specified number of Clocked Periods in a digital exponential decay or charge is equivalent to two (2) Time Constants in an equivalent analogue exponential decay or charge curve.

In other words, when we specify a (say) 100-day recursive filter, at the 100th day, the output of the filter response (from a Step Input) will equal that of two Time Constants (86% of the max amount).

We now have the maths to accurately predict the output of the filter from any known input – no guessing! Thanks, to those other earlier brilliant mathematicians; John Napier[1] who came up with exponentials and natural logarithms[2] and Oliver Heaviside who came up with Singularity Functions[3] and a whole lot more!

We can use this foundational maths to calculate the response to a ramp, and the calculate the error too! This is leaping a forward a little, but there is a whole lot of knowledge about (Feedback) Control Theory[4] that can be applied (in its simplicity) to charting security prices and understanding the errors involved.

Recursive Digital Filters

One way to structure digital filters on a more efficient basis than a Simple Moving Average filter is to use some of the output of a filter and apply it to the input of the same filter. This change makes the filter "recursive" as some of the output reoccurs in the input, making the filter response appear infinite in length.

This family of filters also have the name Infinite Impulse Response (IIR) filters, because the line of maths involved is based on an infinite impulse, which is one of Oliver Heaviside's Singularity Functions.

In this case commonly used for Technical Analysis, this very simple IIR filter has only one stage and takes a (small) percentage of the previous output. The equation for this simple IIR Digital Filter is:

$y_{(t)} = k \cdot x_{(t)} + (1-k) \cdot y_{(t-1)}$ Where k is positive and less than 1.0

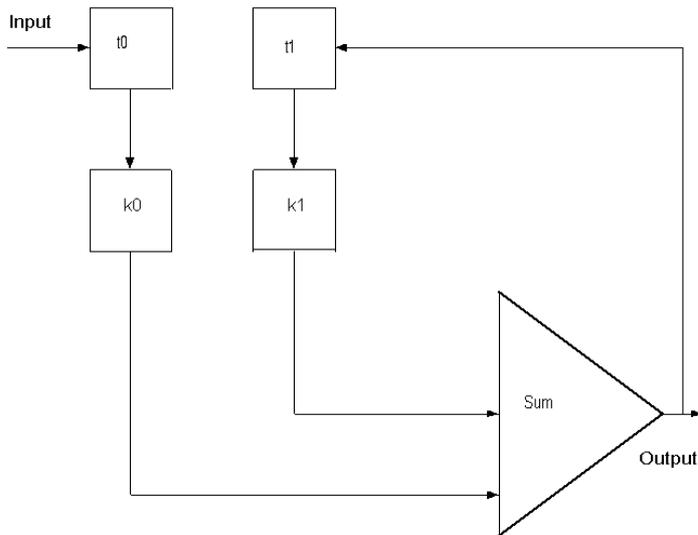---

[1] http://locomat.loria.fr/napier/napier1619construction.pdf
[2] http://en.wikipedia.org/wiki/Natural_logarithm
[3] http://www.engr.sjsu.edu/kghadiri/EE112/Lecture_Notes/Singularity%20function.pdf
[4] http://www.control.utoronto.ca/~francis/dft.pdf

Schematically the drawing of this very simple IIR filter used in Technical Analysis looks like that below.

Input

| t0 | t1 |

| k0 | k1 |

Sum

Output

What is not clearly explained in the picture above is that the boxes t0 and t1 are really clocked Sample and Hold circuits. There is a clocking pulse that causes the input of these boxes to Sample and Hold the input value at the output of their boxes. The inputs can vary all they want but the output remains fixed until the next sample is clocked through.

The two boxes k0 and k1 are "constant" multipliers that put a constant portion of their inputs through to their outputs. The Triangular box is a summing point that adds up the inputs and places this on its output.

In this case the sum of the constants k0 and k1 equals 1 so that if a new value is placed on the input, after several clocking pulses, the output will also equal 1, and the output follows an exponential decay curve to get to this value.

Series1
Series2
Series3
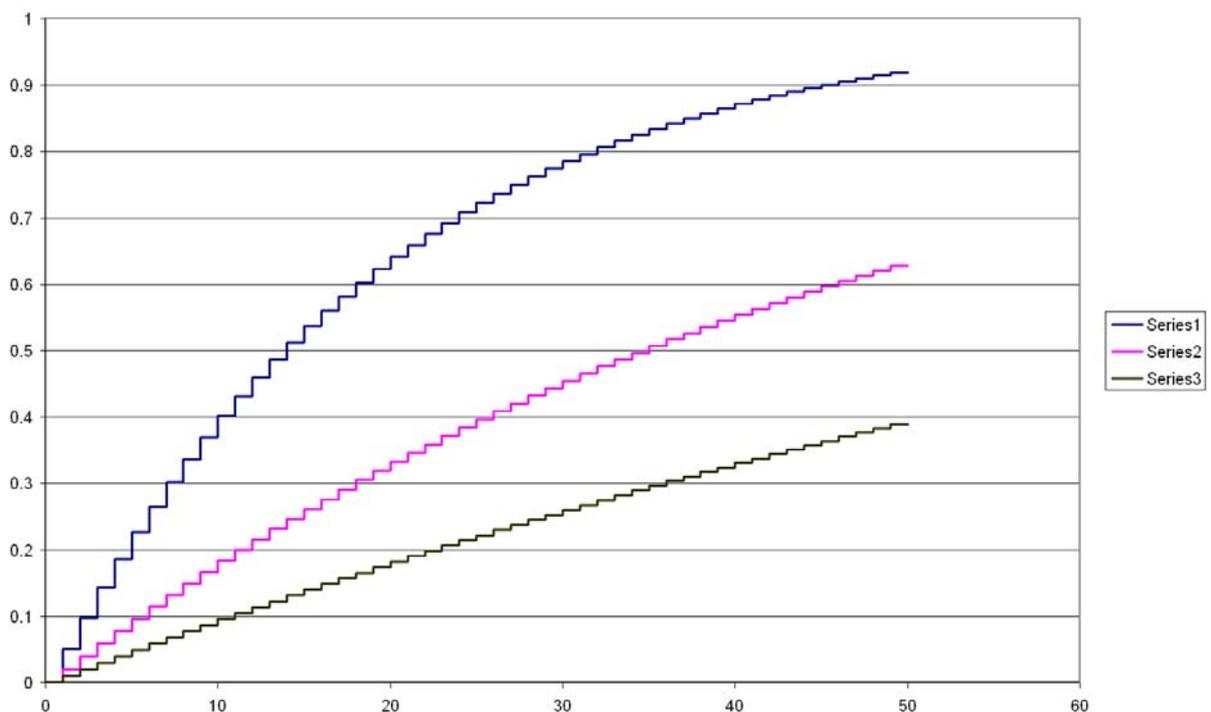
The graph above shows what happens when a Unit Step Singularity Function is fed into the input of the above low (frequency) pass filter. As the clock ticks away, the Sample and Hold values are progressively fed around the block circuit and the output progressively builds towards the unit value of the input.

The rate of time (or count) in clocking pulses before the output settles to insignificant changes is determined by the constants k0 and k1. As we know that k0 + k1 = 1 we only have to specify one constant (k0) and the other one is directly known.

Three example outputs are shown here to show what goes on:

With a 5 % value for k then k0 = 0.05, (k1 = 0.95) then Series 1 (the dark blue) is the first typical exponential transient response.

If the input percentage (k) is dropped to 2% (k0 = 0.02) then Series 2 (the crimson under Series 1) is the expected result.

With k dropped further to 1% (k0 = 0.01) then we have Series 3 (the green line under the other two outputs) is the response.

These outputs all follow stepped exponential time responses. The green steps look to be a linear progression but it has taken nearly 50 clock counts to get to a mere 40% of the final value. The dark blue steps also look (almost) linear up to the 40% point, but that was in just 10 clock counts.

So, with a little feedback we have created a rather simple recursive filter!

Calculating the K Factor

With Technical Trading, the common denominator is periods (usually days), so it is necessary to relate the recursive factor (k) into a Period factor. Fortunately there is a given direct relationship and it is through the formula as follows:

Periods = (2/k)-1

If we chose k = 0.09, this formula converts to 21.2222 Periods, and for k = 0.05, this formula converts to 39.0 Periods and for k = 0.01, this formula converts to 199.0 Periods. So, going backwards, we really want to find out the k factor from the Period and by transposing the formula it becomes:

k = 2/(Periods +1)

So for 11.0 Periods then k = 0.1666666, for 21.0 Periods then k = 0.090909 and for k = 40.0 Periods then k = 0.0487804.

This all appears very simple, but the relationship needs to be tied. Referring back to the graph, it is obvious that the time response is an exponential decay.

What we need to do is to tie the coefficient of a continuous exponential decay with the feedback coefficient of a (first order) digital low pass IIR filter ands then we have a direct correlation between the analogue and digital responses.

So if we have an EMA of 40 periods, (commonly written as EMA40, then the K factor is k = 2 / (Periods − 1) = 2 / (40 −1) = 2 / 39 = 0.051282, so the feedback factor is k1 = 1 - k0 = 1 0.51282 = 0.94872.

So we can make this filter very easily as the output (V out) becomes a recursive equation as follows, where (t+) is the new value after the clocking:

Vout(t+) = Vin * k0 + Vout * K1

Vout(t+) = 0.051282 * Vin + 0.94872 * Vout

Note: The output will be a stepped set of timed values in accordance with the clocking and look similar to the graph as shown in the "Clocked Exponential Decay".

For simplicity almost all EMA charts simply put in the newly clocked values and join them to form a piecewise linear curve that does not have any steps in it. This "linear" representation is fundamentally incorrect and misleading.
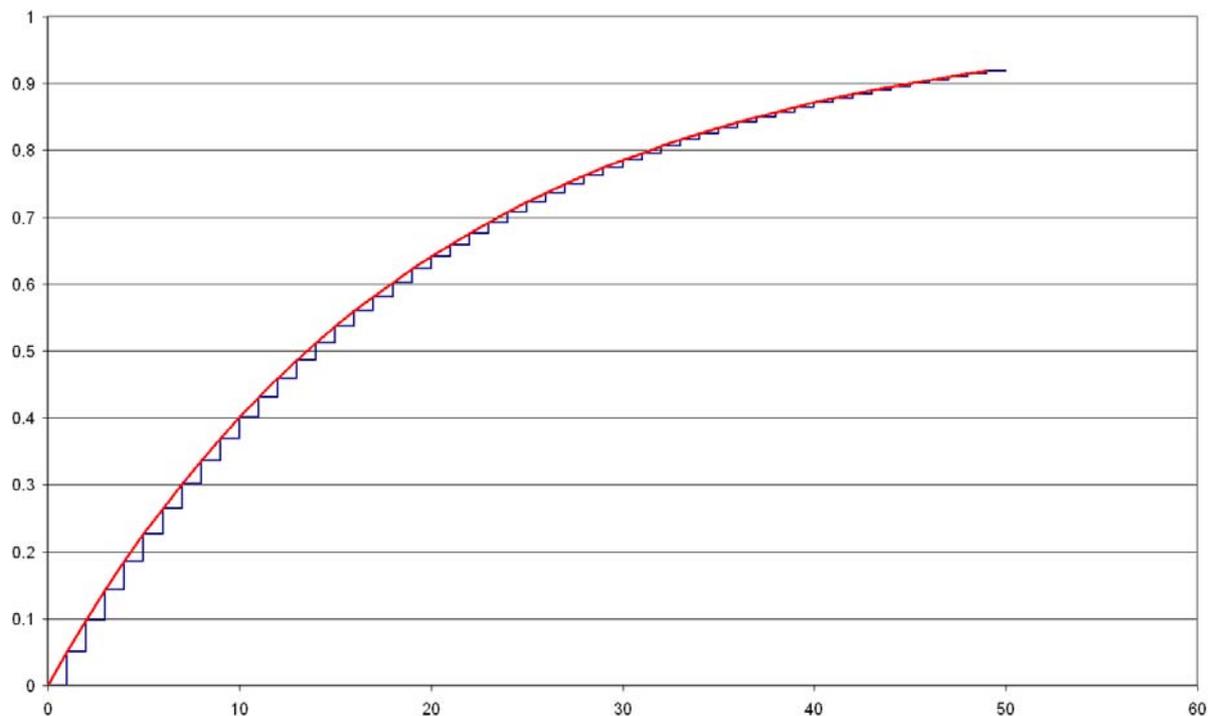
The chart above shows the true EMA as it really is in dark blue steps and as it is usually represented as the red line.