The Gaussian Smoothed Distribution curve function

$$f(x) = \frac{a}{n}\sum_{i=1}^{n}\frac{1}{s\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-d_i}{s}\right)^2}$$

Which simplifies to

$$f(x) = \frac{a}{ns\sqrt{2\pi}}\sum_{i=1}^{n}e^{-\frac{1}{2}\left(\frac{x-d_i}{s}\right)^2}$$

Where:

- $d_1, d_2, d_3, \ldots\ d_n$ are the data values of your single-variable sample
- n is the number of data values in your sample
- s is a smoothing parameter. A good starting value is $\frac{\max(d)-\min(d)}{25}$ ie the range of data values divided by 25. Depending on needs this value can be adjusted.
- a is an amplifying parameter.
  A good starting value is $0.3\ \times\ (x - axis\ range) \times (y - axis\ range)$. This results in the area beneath the curve taking up 30% of the graph paper

Note that $\int_{-\infty}^{\infty} f(x) = a$

Note that each term of the top $\sum$ sum is almost identical to the probability density function of a Normal distribution ϕ(x) but with d_i instead of μ and s instead of σ. So the function sums lots of individual little bell (normal distribution) curves at each data point. As the σ of these bell curves is increased, the bell curves merge into each other creating a 'smoothing' effect, which the function effects by increasing s. Whatever the value of s (ie σ), the area under each mini bell curve is 1, so when added all together the total area remains a constant n.

With a computer, the above function is very easily programmed with a simple 'loop' as part of the function:

```
s=(max(d)-min(d))/25
a=0.3*(max(d)-min(d))*yaxisheight
for (x=min(d) to max(d))
  y=0
  for (i=1 to n)
    y = y+exp(-0.5*(x-di)*(x-di)/(s*s))
  y=y*a/(n*s*sqrt(2*PI))
  plot (x,y)
```

In practice, extending the x-axis a little beyond the minimum and maximum data values is good and allowing s and a to be varied easily (eg with a 'slider' control) in a dynamic fashion enables further insight into the distribution and an optimum view for the data concerned.

Plotting the curve also lends itself to showing more than one data-set on the same graph. Indeed, it gives a good visual way to compare different distributions even of different sizes (n).

Other sources have suggested precise statistical methods for obtaining an optimal value of s, but my personal experience shows the simplified version above gives a good starting point and I like to then vary it dynamically before selecting a compromise between smoothness and detail. The other sources seen also use the sample standard deviation which feels uncomfortable for samples from a non-normal, possibly bi-modal skewed distribution.