Estimation Uncertainty of the Coefficient of Variation

The purpose of this white paper is to examine the uncertainty in the estimate of the Coefficient of Variation.

The Coefficient of Variation

The coefficient of variation (*CV*; or relative standard deviation) of any variate c (for example, concentration) defined by a sample population $\{c_i \mid i = 1..n\}$ is expressed as

$$CV = s_c / \bar{c} \tag{1}$$

where $\bar{c}$ is the sample mean

$$\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{2}$$

and s_c is the (unbiased) sample standard deviation

$$s_c = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(c_i - \bar{c}\right)^2} \tag{3}$$

It is important to note that since the sample mean and sample standard deviation are defined over a *finite* population, they themselves are random variates with associated uncertainties in their estimates.

Familiar to most is the uncertainty in the sample mean, given simply by

$$\delta(\bar{c}) = s_c / \sqrt{n} \tag{4}$$

Certainly less familiar is the uncertainty in the sample standard deviation given by

$$\delta(s_c) = s_c\sqrt{\left(\frac{n-1}{2}\right)\cdot\left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)^2 - 1} \tag{5}$$

where the symbol Γ represents the Gamma function (related to the factorial function for integer values of the argument i.e. $\Gamma(n) = (n-1)!$). Fortunately, for values of $n > 12$ expression (5) can be approximated to better than 1% by the much simpler expression

$$\delta(s_c) = s_c \Big/ \sqrt{2(n-1)} \tag{6}$$

Uncertainty in the Coefficient of Variation

From expression (1), we now see that the *CV* is defined in terms of the ratio of two random variates, each with an associated uncertainty. Therefore, the uncertainty in the *CV* is computable using the propagation-of-errors formula. In the case here, the two random variates $\bar{c}$ and s_c are uncorrelated, so the propagation-of-errors formula may be written

$$\frac{\delta(CV)}{CV} = \sqrt{\left(\frac{\delta(\bar{c})}{\bar{c}}\right)^2 + \left(\frac{\delta(s_c)}{s_c}\right)^2} \tag{7}$$

Substituting the uncertainties from expressions (4) and (6) yields

$$\frac{\delta(CV)}{CV} = \sqrt{\frac{1}{n}\left(\frac{s_c}{\bar{c}}\right)^2 + \frac{1}{2(n-1)}} \tag{8}$$

Recognizing the definition of CV in the first term under the radical, we further simplify to obtain

$$\frac{\delta(CV)}{CV} = \frac{1}{\sqrt{n}}\sqrt{CV^2 + \frac{1}{2}\left(\frac{n}{n-1}\right)} \tag{9}$$

Behavior of the Uncertainty of the CV Estimate

For modestly large values of the number of samples, say 10 or more, the relative uncertainty of the CV can be further simplified as

$$\frac{\delta(CV)}{CV} \approx \frac{1}{\sqrt{n}}\sqrt{CV^2 + \frac{1}{2}} \tag{10}$$

If we consider that for most applications it is desirable that the CV be small, say 20% or less, we can further simplify this expression to obtain

$$\frac{\delta(CV)}{CV} \approx \frac{1}{\sqrt{2n}} \tag{11}$$

For example, from this expression we see that if the number of samples is 25 then the relative uncertainty in the CV is 14.1%. On the other hand, if it is desired to have a relative CV error of 10%, then the number of samples must be 50.