
Reducing the Effects of Noise in a Data Acquisition System by Averaging

E. B. Loewenstein

Introduction

There are two important statistics involved in the investigation of signal noise. The first is the mean, or average, or expected value of a variable. This quantity is often mathematically denoted $E[X]$, where X is a sample of the noise in question and $E[\cdot]$ is called the *expectation* (average value) of the quantity inside the brackets. This parameter is usually the DC voltage we are trying to measure, to which noise is being added. The second statistic is the standard deviation of the noise. This is computed by subtracting the square of the mean from $E[X^2]$ and taking the square root. The standard deviation is a measure of the magnitude of the energy of whatever AC signal is present (just noise, we hope, in the case of a DC measurement) and is independent of whatever DC signal is present. It should be noted that the standard deviation is not quite the same as rms, which is simply the square root of $E[X^2]$. True rms includes the DC with the standard deviation, although for signals with 0 mean it is the same as standard deviation.

We start with the assumption that we have a group of samples that are *i.i.d.* (independent and identically distributed). This statistical property is really two properties in one, but they are lumped into one term because they often occur at the same time.

The first property is that all data samples are independent from each other, which means that no information about any one sample can be inferred from knowledge of any other. This assumption is not always safe, because the samples usually are not completely uncorrelated, especially when there is (possibly inherent) filtering somewhere in the signal path. With lowpass filtering, samples near each other (in time) tend to be correlated; the lower the cutoff, the more that adjacent samples tend to be correlated and the longer the time periods over which the correlation occurs. If the cutoff frequency is not much below the Nyquist rate (half the sample rate), then the correlation among samples is not significant and they can be considered independent. If the cutoff frequency is significantly less than the Nyquist rate, then averaging is not as effective as it is ideally.

The other property, identical distribution, means that the statistical distribution of each of the samples is the same. Specifically, the means of all the samples are the same, as are the standard deviations. This assumption is convenient, because in calculations we can now use the same statistics to describe each of the samples. As with independence, it is not always a reasonable assumption, because the character of noise (additive interference) is often time varying. For instance, digital logic in a data acquisition system behaves differently at the beginning of a data acquisition sequence than at other times. Hence, interference coupling from the logic activity at the beginning might produce a higher standard deviation in the samples at the beginning than it does in the rest of the samples, or it might even produce different means at different times. In general, the less the statistics of the noise or interference vary with time, the more predictable averaging is.

Derivations

Given that we have i.i.d. samples, we can talk about the statistics of any sample by using statistics that apply to all the samples equally. If we have a sequence of n samples (X_i), let the mean of any sample be denoted μ , where, by definition

$$\mu = E[X_i] \quad (1)$$

Because the samples are identically distributed, they all have the same mean and standard deviation, so i can refer to any one of the samples. Let the standard deviation be denoted σ , where, by definition

$$\sigma^2 = E[X_i^2] - \mu^2 \quad (2)$$

and where, again, i refers to any one of the samples.

We consider the statistics of a simple average, which we call X_{avg} , taken over the n samples (X_i). The formula for this average is simply

$$X_{avg} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3)$$

Denote the mean of the average μ_{avg} and the standard deviation of the average σ_{avg} . Then

$$\mu_{avg} = E[X_{avg}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu \quad (4)$$

and

$$\begin{aligned} \sigma_{avg}^2 &= E[X_{avg}^2] - \mu_{avg}^2 = E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2\right] - \mu^2 \\ &= E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right)\left(\frac{1}{n} \sum_{j=1}^n X_j\right)\right] - \mu^2 \\ &= E\left[\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n X_i X_j\right] - \mu^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E[X_i X_j] - \mu^2 \end{aligned}$$

Now we split up the first term above into two terms, and we use the assumption of independence, which gives us that $E[X_i X_j] = E[X_i]E[X_j] = \mu\mu = \mu^2$ whenever $i \neq j$. Thus, continuing

$$\begin{aligned}
 \sigma_{avg}^2 &= \frac{1}{n^2} \sum_{i=1}^n E[X_i^2] + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n E[X_i X_j] - \mu^2 \\
 &= \frac{1}{n^2} \sum_{i=1}^n (\sigma^2 + \mu^2) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n \mu^2 - \mu^2 \\
 &= \frac{1}{n} (\sigma^2 + \mu^2) + \frac{n-1}{n} \mu^2 - \mu^2 = \frac{1}{n} \sigma^2 \quad (5)
 \end{aligned}$$

Hence

$$\sigma_{avg} = \sigma \sqrt{\frac{1}{n}} \quad (6)$$

as expected.

Discussion

We see from equation (6) that the standard deviation of the *average* of the samples is reduced from the standard deviation of each individual sample by a factor of the square root of the reciprocal of the number of samples. Furthermore, equation (4) shows that the mean of this average is the same as the mean of each sample, so the DC content of the data is preserved. In general, increasing by a factor of n the number of samples that are averaged results in a decrease by a factor of the square root of n in the standard deviation of the average. For example, suppose that the σ of some acquired data tends to be about 1 LSB. According to equation (6), the σ_{avg} resulting from an average of four points is 0.5 LSB, so the measured noise is cut in half. If 100 points are taken and averaged, the resulting σ_{avg} is 0.1 LSB.

Although the goal of averaging is to reduce noise, there is actually some benefit to having system noise in the first place. In fact, noise added to a signal going into an analog-to-digital converter (ADC) can actually *increase* the resolution of the converter beyond the number of bits of resolution it provides. This increase can be achieved accurately only when there is a certain amount of noise ($\sigma \geq 0.5$ LSB) present in the signal, which is the principal of dither. Under this condition of sufficient noise, the mean of the data that an ADC returns is very close to the actual mean of the incoming analog signal. If this condition is not met, the mean of the digital data may be as much as ± 0.5 LSB from the actual mean (notwithstanding converter nonlinearities or other imperfections). Hence if a converter has good linearity and sufficient dither is present, its resolution can be extended by averaging the data it produces. As above, if a system adds 1 LSB of noise to the data but 100 samples are averaged, then the measured noise is reduced to 0.1 LSB and the effective resolution is increased by several bits.

Finally, it should be noted that a simple average as described above is not the only mechanism for “averaging out” noise. Other examples are weighted averages, moving averages, and moving-weighted averages. Specifically, a digital filter is a moving-weighted average. The advantage of a digital filter over simple averaging is that it can reject regions of the spectrum where noise is known to prevail without reducing information in other spectral regions (in this case the samples are not completely independent). However, the actual reduction in noise is somewhat harder to predict under such circumstances.

Conclusion

Underlying the application of averaging or filtering is a trade-off between the degree of certainty achieved and the number of samples that must be taken (and the time it takes to obtain them). When samples are independent and identically distributed, averaging a collection of samples reduces measurement uncertainty by a predictable amount. If σ is the amount of rms noise in a set of n i.i.d. samples, the rms noise of the average taken over the samples is

$$\sigma_{avg} = \sigma \sqrt{\frac{1}{n}}$$

