1.7 Uncertainty

Kline & McClintock (1953) defined uncertainty as “A possible value the error might have” and this is the definition we still work with today. In most engineering experiments it is simply not practical to repeat an experiment enough times to provide uncertainty statistics. Even if multiple trials are feasible it generally is only possible with the same measurement equipment/sensors. Thus in the global sense almost all experiments are single-sample. How might we ascribe the uncertainty to a measurement? If there are multiple independent measurements one possibility is in terms of the standard deviation, or perhaps the range as defined by the maximum value minus the minimum value. Kline & McClintock laid the foundation for single-sample uncertainty analysis and proposed that for single-sample uncertainty, an uncertainty, or confidence interval approach, be taken.

1.7.1 Basics

Consider a suite of variables under measurement which we will denote by $X_i$. Let $\delta X_i$ be the uncertainty in a measured value of $X_i$ and let $\hat{X}_i$ be the measured value (or mean of $N$ measured values if the sample is multiple). We assume that the best estimate of $X_i$ is $\hat{X}_i$ and we can represent the uncertainty in our measurement as

$$X_i = \hat{X}_i \pm \delta X_i$$ \hspace{1cm} (1.150)

Typically the above is stated to occur with some given probability, or odds. Hence

$$P[|X_i - \hat{X}_i| > \delta X_i] = \alpha$$ \hspace{1cm} (1.151)

where $\alpha$ is the $(1 - \alpha) \times 100\%$ confidence level (or $\frac{1}{\alpha}$: 1 odds).

It is generally assumed that uncertainty statistics will be Gaussian although this assumption should be kept in mind. For a single sample the Gaussian assumption leads to choosing $z_{\alpha/2}\sigma$ as the bounds on the confidence interval. Taking $\alpha=0.05$, as is commonly

© 2017 Edwin A. Cowen
done, leads to the 95% uncertainty interval (20:1 odds) and

\[ \delta X_i = 1.96\sigma \]  

(1.152)

where \( \sigma \) is the standard deviation of the possible measurements. If the experiment is multiple sample, one interpretation is based on random error analysis, following the single-sample experiment

\[ \delta X_i = \frac{t_{(N-1)\alpha} s}{\sqrt{N}} \]  

(1.153)

where \( s \) is the sample standard deviation and \( t \) is Student’s \( t \)-distribution with \( N - 1 \) degrees of freedom. It may also represent the bias limit or the overall uncertainty, a combination of the bias limit and the random error.

An important question is how does error in measured individual variables combine to give the total error in some desired result, which we will denote \( R \). In general \( R \) is calculated in some way from the data (measured variables), \( \hat{X}_i \). Thus we can write

\[ R = R(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_M) \]  

(1.154)

where there are \( M \) variables that define \( R \). Now, \( R \) can be Taylor series expanded in the \( \hat{X}_i \)'s. If the variation in the \( \hat{X}_i \) are small then a linear approximation (truncation of all non-linear terms in the Taylor series expansion) provides a reasonable approximation of the propagation of the uncertainty into the result from uncertainty in the data. Thus we can write

\[ \delta R_{X_i} = \frac{\partial R}{\partial X_i} \delta \hat{X}_i \]  

(1.155)

The partial derivative terms are known as sensitivity coefficients. Kline & McClintock demonstrated that the method to combine the \( \delta R_{X_i} \) optimally is the root-sum-square (RSS) technique given by

\[ \delta R = \left[ \sum_{i=1}^{M} \left( \frac{\partial R}{\partial X_i} \delta \hat{X}_i \right)^2 \right]^{1/2} \]  

(1.156)

The fundamental assumptions in the above are

- The measurements are independent

© 2017 Edwin A. Cowen
• The probability density function of the $\hat{X}_i$ are Gaussian

• The uncertainty interval in each measurement is the same

From Eq 1.156 it is clear that small terms have little effect. Hence Eq 1.156 can be used to determine which data in an experiment most impact the uncertainty in a result and hence where energy and money should be spent to maximally reduce uncertainty.

### 1.7.2 Single-Sample Uncertainty in Stochastic Measurements

Most of the data reported in fluid mechanics is the result of the ensemble averaging process. The determination of a mean value or an RMS value is conducted only once, but from a large data set with $N$ samples in it. Moffat (1988) presents a formal methodology based on single and multiple sample analysis to determine the uncertainty in these types of measurements. He proposes splitting the data into a fixed or bias error and a variable or random error. The bias error is the fixed or slowly varying error that is deterministic, but generally not separable from the measurement (e.g., calibration error). Essentially bias errors have a fixed mean and zero variance. Random errors arise from many sources (e.g., thermal noise, electronic noise, etc.) and are characterized as having zero mean but non-zero variance. The two error types, in general, are best considered separately since if the data is differenced the bias errors are eliminated to leading order, leaving only the random error.

In single-sample experiments one option for determining the bias and random errors is to perform an auxiliary experiment, typically with 30 replicates, to determine these components. This is described in Moffat. An alternative, slightly less formal method, is to estimate the bias error in the variables at the 95% level (which may be done conservatively as a ‘worst case’ estimate, for example the worst case error in the measurement in length using a millimeter graded ruler is 1 mm) and then use the RSS technique to determine the bias error in the reported results. This leaves the issue of random error.
1.7.3 Bootstrap Uncertainty Analysis

The bootstrap algorithm is surprisingly simple yet surprisingly powerful. Due to its heavily computational nature it was not even conceived of until well into the computer age. It was first proposed by Bradley Efron in 1977 and has spread widely through all literature – wherever statistics has previously been used. For a thorough treatment see Efron & Tibshirani (1993) in the reference list. Diaconis & Efron (1983) (Computer-Intensive Methods in Statistics, *Scientific American* 248, 116-130, available at \handout on the course website), gives widely varied examples of the application of bootstrap analysis.

The bootstrap algorithm is incredibly simple. Given a set of $N$ measurements we often like to calculate statistics from the measurements - the mean and the variance being some of our favorites but there are myriad other possibilities. The bootstrap is a method to directly estimate the probability density function of any statistic we can imagine that can be calculated from the measurements.

For example, consider the mean. Normally we would calculate the mean and have a single-sample estimate of the mean. The bootstrap algorithm gives us a way to replicate the estimate into a multiple-sample estimate and hence to directly estimate the probability density function for the mean of the measurement. This is done by selecting a set of $N$ samples from the original $N$ with replacement. This means that any individual measurement may be repeated multiple times within a given set. Now we can calculate our statistic, the mean in this case, from the sample. If we repeat this process $M$ times we can now estimate the probability density function for the statistic under consideration - the mean.

As an example consider the bootstrap replication (at $M=1000$ replicates) of the mean value of the number of heads that appeared on a toss of a coin 8 times by 15 people (same example as in next section). The probability density function would look something like the distribution shown on the top of the next page giving us the sense that the statistic

© 2017 Edwin A. Cowen
under study is Gaussian-like (it is binomial as you may recall) and that an uncertainty interval can be defined such that the area under the probability density function is 0.95.

1.7.4 Bootstrap Percentile Technique

The uncertainty interval in a statistic can be directly estimated at a given level from the ordered statistic (e.g., sorted from lowest to highest value of the statistic). Efron & Tibshirani (1993) suggest $M=1000$ replications to be sufficient for the 95% uncertainty interval. Thus for $M=1000$ the bootstrap replicates of the statistic are determined and then sorted lowest to highest. The 25th and the 975th values of the ordered statistic are then taken as the lower and upper bounds, respectively, of the 95% uncertainty interval.

As an example consider if 15 members of the class each flipped a coin 8 times as a way of estimating the number of times a head would occur for 8 coin tosses. If our result was:

\[
\begin{array}{cccccccccccc}
2 & 2 & 7 & 4 & 4 & 3 & 6 & 3 & 4 & 4 & 3 & 4 & 7 & 4
\end{array}
\] (1.157)

we would find the mean is exactly 4, but what is the uncertainty in our estimate? If we turn to the bootstrap we can calculate an ordered statistic (shown at the top of the next page) where the 25th statistic is 3.27 and the 975th is 4.80. Our best estimate for the number of heads is still the mean value that we measured, 4, and hence we report the uncertainty interval as 4.00 +0.80, - 0.73. Note that the 1000 drawn bootstrap replicates

© 2017 Edwin A. Cowen
are random so the exact uncertainty interval is a random process itself (and the bootstrap could be used to look at the 95% uncertainty interval on the bounds themselves, which is left as an exercise for the student!)

The bootstrap is easily programmed in a language such as C or FORTRAN, however, we are lucky – it is already in Matlab as the function `bootstrp`. The Matlab code to perform the above analysis is simply:

```matlab
x=[2 2 7 4 4 3 6 3 4 4 3 4 3 7 4];

y=sort(bootstrp(1000,'mean',x));
mean(x)
UL_lo=y(25)-mean(x)
UL_hi=y(975)-mean(x)

plot(y)
```

How does our finding of 4.00 +0.80, - 0.73 from the bootstrap analysis compare to an assumption of normal statistics? Since our number of trials is low we turn to Students’ t distribution. We have $N - 1 = 14$ degrees of freedom and a sample standard deviation
of $s = 1.5584$ where we have used the unbiased estimate of the standard deviation (e.g., normalized by $N - 1$). Thus our Students' t based uncertainty interval occurs at
\[ \mu - t_{14;0.025}s/\sqrt{N} \] and \[ \mu + t_{14;0.975}s/\sqrt{N} \] which gives $4 +/-0.86$. Not surprisingly in this case a Gaussian assumption is pretty good. Note in matlab $t_{14;0.025}$ can be evaluated as $tinv(0.025, 14)$.

As a further example of the bootstrap and the analysis of the bias and total errors I have handed out the uncertainty analysis from my thesis.