

# Measurement Uncertainty

Barton L. Smith and Douglas R. Neal

September 7, 2025



# Contents

<b>1 Introductory Example</b>	<b>9</b>
<b>2 Introduction</b>	<b>17</b>
2.1 Errors . . . . .	21
2.2 Uncertainty . . . . .	26
2.3 Nomenclature . . . . .	27
2.4 Propagation . . . . .	30
2.4.1 Taylor Series Method . . . . .	30
2.4.2 Monte Carlo Method . . . . .	32
<b>3 Probability</b>	<b>35</b>
3.1 Random Variables . . . . .	35
3.2 Gaussian . . . . .	39
3.3 Uniform . . . . .	42
3.4 Periodic . . . . .	42
3.4.1 Sinusoidal . . . . .	42
3.4.2 Sawtooth . . . . .	45
3.5 Central Limit Theorem . . . . .	47
<b>4 Single Variable</b>	<b>51</b>
4.1 Random Uncertainty . . . . .	52
4.1.1 Random Uncertainty of the Mean for Independent Data . . . . .	54
4.1.2 Random Uncertainty of the Mean of Small Samples . . . . .	61
4.1.3 Convergence of Statistics . . . . .	65
4.1.4 Outlier Detection . . . . .	66
4.1.5 Random Uncertainty of the Mean for Correlated Data . . . . .	67

4.1.6	Phase and Conditional Averaging . . . . .	71
4.2	Bias Uncertainty . . . . .	73
4.2.1	Uncertainty From Reading Analog Devices . . . . .	74
4.2.2	Periodic Inputs . . . . .	74
4.2.3	Instrument Calibration . . . . .	76
4.2.4	Using Instrument Specifications For Uncertainties . . . . .	80
4.2.5	Examples of Use of Instrument Specification . . . . .	85
4.2.6	Uncertainty From Material Property Data . . . . .	88
4.2.7	Uncertainty of Coefficients of a Polynomial Fit . . . . .	88
4.2.8	Uncertainties in Line Fit Parameters . . . . .	91
4.3	Dynamic Range . . . . .	93
4.4	Effects of Spatial and Temporal Resolution . . . . .	95
4.5	Total Uncertainty . . . . .	96
<b>5</b>	<b>Uncertainty Propagation</b>	<b>103</b>
5.1	TSM . . . . .	105
5.2	MCM . . . . .	112
5.3	Planning . . . . .	131
<b>6</b>	<b>Epilogue</b>	<b>139</b>
<b>A</b>	<b>Appendices</b>	<b>145</b>
<b>A</b>	<b>Statistics Tables</b>	<b>149</b>

# Preface

This text is based on the graduate Measurements course notes from BLS at Utah State University. It is intended as a textbook for graduate students in engineering or physics or as a reference text.

If one collects several texts on measurement uncertainty, it quickly becomes clear that the measurement uncertainty community lacks a consistent language or nomenclature. This can be confusing to a new reader who may view measurement uncertainty as a very old and settled topic. The lack of consensus on nomenclature likely stems from the absence of a clear winner—a paradigm and nomenclature set sufficiently superior to be adopted by everyone. It is our hope that this book will improve on previous texts.

In addition, this text will attempt to move beyond traditional thinking about measurement uncertainty that is limited to point measurements and uncertainties that are either fixed or linearly related to the reading. With a special focus on camera-based measurements and their unique issues, it is our ambition to write the first modern measurement uncertainty text.

The authors wish to acknowledge a debt to Hugh Coleman and Glenn Steele and their 3rd edition text [1], which made progress in the directions that we hope to continue; towards a richer and more realistic uncertainty analysis that considers material property uncertainties as well as the positive and negative impacts of correlated uncertainties. This text was later revised into the 4th edition [2], which added some small but significant changes.

Additionally, the authors would like to thank many individuals who gave specific input. Steve Beresh and Randy Hurd were kind enough to provide a very comprehensive review of the text. Geordie Richards guided us through several issues with respect to statistics. Alan Nathan reviewed

some of the baseball examples and Bill Zwolinski engaged in many fruitful conversation about instrument specifications. Finally, BLS would have never been interested in this topic without the excellent mentorship of Dr. Greg Swift of Los Alamos National Laboratories.

# Nomenclature

$\bar{\delta}_X$	The mean total error on $X$
$\bar{X}$	The sample average of $X$
$\delta_X$	The total error on $X$
$\mu_X$	The true (parent) mean of $X$
$\sigma_X$	The true (parent) standard deviation of $X$
$s_{\bar{X}}$	The sample standard deviation of the mean of $X$
$s_X$	The sample standard deviation of $X$
$u_{X_i}$	The standard uncertainty of $X$ from the $i$ th known error source
$u_{X_r}$	The standard random uncertainty of $X$
$X$	The measured value of $X_{\text{true}}$
$X_{\text{true}}$	The unknown true value being measured



# Chapter 1

## Introductory Example

While the authors derive a great deal of enjoyment from the topic of Measurement Uncertainty, we acknowledge that, at the start, many readers may not feel the same way. The topic sounds dry. However, it is all around us. Understanding measurement uncertainty can help us better understand much of the information we are presented with daily, from political poll results, to sports metrics, to the readouts in our automobiles.

The following example will be used to illustrate and motivate many of the important topics of this text. The discussion will be about concepts and rigorous definitions will be provided in later chapters. Whenever possible, a reference will be made to the chapter and section where each topic is discussed in more detail.



Figure 1.1: Cars traveling at highway speed with close spacing.

Imagine you are in a line of 20 cars that are unwisely traveling bumper to bumper traveling on the Autobahn. Due to the fixed spacing between the cars, we know that each car is traveling at nearly the same speed. Each

driver reports the speedometer reading (Fig. 1.2) at the same time that a radar is used to determine the speed of the lead car. The radar reads 129.5 kph. The values reported by the 20 drivers are shown in Fig. 1.3.



Figure 1.2: Automobile speedometer reading 129 kph.

Each reading is unique. Most are near the radar reading of 129.5 kph with about half above and half below that value. This process is repeated with the same 20 cars. In all, 40 trials are held. The radar reading is similar for all 40 trials, but the speedometer reading changes in each car for each trial. The results are shown, superimposed, in Fig. 1.4.

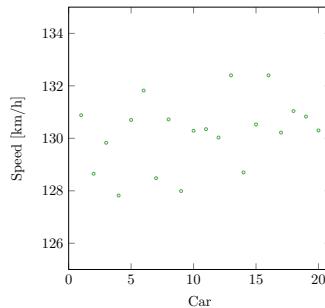


Figure 1.3: The speeds reported by all twenty cars for the first trial.

A histogram of one trial (#6) is shown in Fig 1.5. A histogram plot shows how often a range of values (called a bin) occurs. This makes it clear that the values near the radar reading happen more frequently than speeds far above or below that value. A histogram shows us the *distribution* of values, or how often values inside a bin occur. How often a certain value occurs is often described as its *probability*, and distributions of an infinite number of samples are called a *probability density*. A histogram of a set of samples provides information about the distribution from which the samples were drawn. While there are insufficient data points in Fig 1.5 to

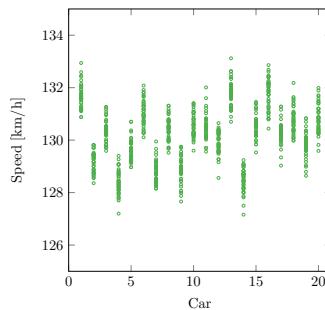


Figure 1.4: The speeds reported by all twenty cars for 20 trials.

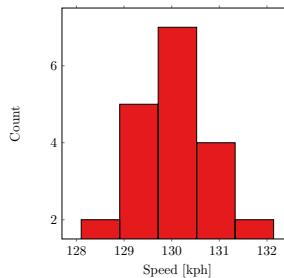


Figure 1.5: Histogram of trial 6.

prove it, these values are drawn from the *Gaussian* probability density (cf. Sec. 3.2), which is the most useful probability distribution for measurement uncertainty. The width of the distribution is a measure of the velocity's uncertainty.

Examining the 40 trials, it is clear that the reported speeds are clustered closer to the radar reading, with more readings nearer to 129.5 kph than farther from it. This is called a *central tendency* (cf. Sec. 3.2), which is one of the features of a Gaussian probability density, and is clearly present for all cases. It also appears that the center of the distribution for each car is different. We note that some cars that had a particularly high value in Fig. 1.3 now have values centered around a higher value (e.g. car 6 or 13). This suggests that these cars have a speedometer that tends to read high. Others (e.g., car 4) tend to read low.

This tendency seems to be small compared to how much the speed

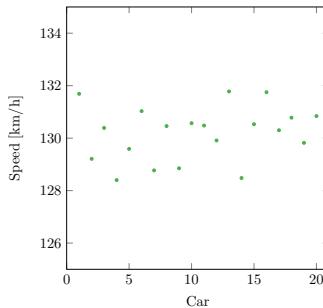


Figure 1.6: The average reading of the car speed for 40 trials.

measurement varies from one trial to the next. This random variation from one trial to the next can be removed by averaging the 40 trials to provide a sense of how much the speedometers vary from one car to the next. This is shown in Fig. 1.6 and looks similar, but not the same, as the result from the first trial.

Here is what is occurring: Each speedometer is calibrated at the factory and is only as accurate as the device that it is calibrated against. Some fixed *error* (difference between the reading and truth, cf. Chapter 2) remains and the error is unique for each car.

In addition to this fixed error for each car, there is an additional error that is different for each trial. This may be due to random variation of the reading (imagine the speedometer needle fluctuating) or random variation in the actual car speed. As demonstrated, we can filter this random error out (to some imperfect extent, cf. Sec. 4.1.1) through averaging. However, the fixed error from the factory remains.

The size of the errors due to the random variation of the reading may be estimated in many ways. In this text as well as most others, the standard deviation of the readings is used to estimate their probability density and the width of the distribution will become an expression of the *random uncertainty* of this measurement (cf. Sec. 4.1). In other words, uncertainty is expressed as a range of values that the true measurement *probably* lies within. It is much less useful from an analysis standpoint to attempt to place bounds on the errors.

While this is not typically possible in an experiment, the width of the distribution of fixed errors can be visualized by a histogram (Fig. 1.7)

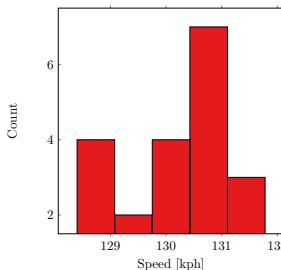


Figure 1.7: Histogram of the mean of the 20 cars.

and estimated by computing the standard deviation of the 20 averaged results. This provides information about the accuracy of the speedometer calibration standard and will be described as an *uncertainty* (cf. Sec. 4.2).

We are able to estimate the random variation from one trial to the next by computing statistics of the various trials. By averaging the trials, we are able to estimate the fixed variation from one car to the next. The statistics of this variation are an expression of the systematic uncertainty.

The true speed of the cars in this example is 130 kph which means the radar device has an error of 0.5 kph. This error remains fixed in all measurements, providing no ability to assess an uncertainty due to this error source. The radar has a specified accuracy of 1 kph, meaning the error is to be expected. It is common for systematic uncertainties to be assessed through instrument specifications rather than from the data itself.

Consider another scenario where the cars were manufactured on three different days. In Fig. 1.8 a-c, the first trial, all 40 trials and the average of the 40 trials are shown. In this case, it is clear that the fixed error from calibrating the speedometers was different on each manufacturing day. In this case, we have only 3 unique data points to assess the variation of the calibration error. This is an example of samples that are not independent of one another and is the topic of Section 4.1.5. Note how averaging reveals the nature of the calibration errors in Fig. 1.8 c and that these are not discernible from the single trial of Fig. 1.8 a.

In a third scenario, there is a “drift” in the factory calibration of the speedometers (Fig. 1.9 a-c). In this case, we have “phase averaged” the results by averaging over the 40 trials to reveal the period of the drift.

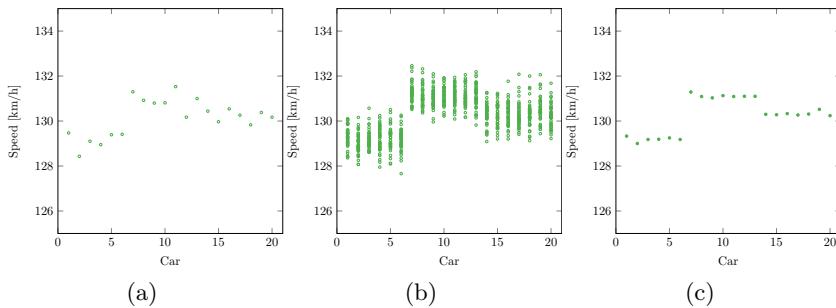


Figure 1.8: (a) One trial, (b) 20 trials and (c) the mean of the 20 trials for a scenario where the 20 cars were manufactured on 3 separate days during which the speedometer calibration error was fixed.

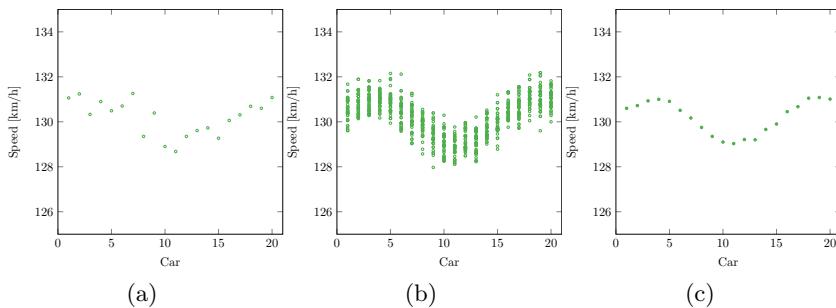


Figure 1.9: (a) One trial, (b) 20 trials and (c) the mean of the 20 trials for a scenario where the 20 cars where the calibration error of the speedometers drifted during the manufacturing process.

This assumes that car 2 was manufactured before 1 and after 3, etc. If the order of manufacturing was not known and the 20 cars were randomized, the random trial-to-trial variations would still be distributed around 3 distinct speedometer errors.

In each of the above examples, the errors, whether they were from one trial to the next or the speedometer calibration error, seemed to have a central tendency. However, this is not always the case. To illustrate a case with a different distribution of errors, consider a new scenario that still involves car speedometers and a radar. In this case, the radar is calibrated against a source that has an accuracy of 0.01 kph and reads out to two digits past the decimal. The car speedometer has a resolution of 1 kph. As before, on each lap, the driver of the car reports the speedometer reading. One thousand laps are recorded over a very boring Saturday afternoon. Since the radar is significantly more accurate and resolved than the speedometer, it may be used as the “ground truth” and the error in the speedometer  $i^{th}$  reading can be computed as

$$\epsilon_i = V_{r_i} - V_{s_i}, \quad (1.1)$$

where  $\epsilon$  is the reading error,  $V_r$  is the radar velocity reading and  $V_s$  is the speedometer reading. In this case, the bulk of the error will come from the speedometer’s lack of resolution. The distribution of the errors is shown in Fig. 1.10. Note that the errors are between 0 and 1 and that any value is as likely as any other value. This is the nature of a *uniform distribution*, discussed in Sec. 3.3. A uniform distribution of errors is a common feature of errors that stem from digitization, or the process of converting a continuous quantity into a discrete, digital one.

Through these examples, we have been able to demonstrate concepts of errors and the distribution of errors. We have discussed how measurement uncertainty can be thought of as how wide the errors are distributed as well as different distribution shapes. We have shown that repeated measurements can provide information that leads to assessment of random uncertainty, but that repeated measurements may not reveal anything about other uncertainties.

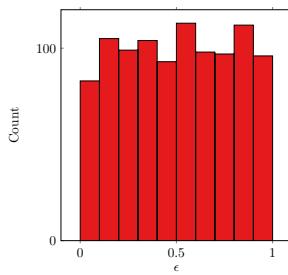


Figure 1.10: Histogram (distribution) of the speedometer digitization error over 1000 laps.

## Chapter 2

# Introduction: Measurement Errors and Uncertainty

This text seeks to provide readers with the ability to assess the range in which the true result of a measurement lies to a desired probability. This is a definition of *uncertainty*. Think of uncertainty as an assessment of the extent to which the result is to be believed.

Measurement systems do not return the true value because they inevitably suffer from *errors*, which are differences between the true values and the measured values. It is a common impulse to seek to eliminate measurement errors, but attempting to do so will invariably lead to unacceptable costs in terms of resources and time. A better approach is to accept the presence of errors and to quantify their impact. It could be, for a specific situation, that the uncertainty of a measurement is something we are not concerned about at all.

The distinction between error and uncertainty is illustrated in Fig. 2.1. An error is the signed difference between a measured value and the (generally unknown) true value. This is an important concept, but since the true value is unknown, we cannot directly analyze errors. Instead, we concern ourselves with uncertainty, which is an estimate of the magnitude of the error, and it is commonly given as an interval with an estimated probability that the true value lies within that interval [3].

A single measurement often suffers from many error sources. Uncertainty analysis seeks to assess the uncertainty of a measurement based

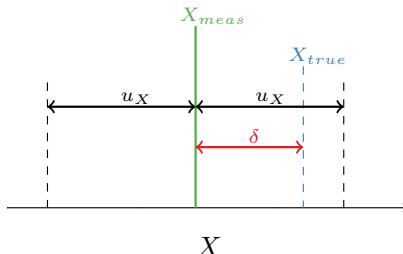


Figure 2.1: An uncertainty,  $u$ , is an interval that probably contains the actual value of an error of unknown sign and magnitude. Red indicates an error, blue is the true value, and green represents a measurement subject to error.

on knowledge of how the various error sources behave, e.g. do they vary rapidly, slowly, in sync with each other, etc. Generally, the value of an error is unknown. If it were known we would simply correct it!. However, even with the unknown value, the probable range error magnitudes may be assessed.

There are numerous motivations for running experiments, but regardless of the need for experimental data, they are less valuable without an *uncertainty analysis*. As engineers and scientists, we often have perfectionist tendencies, but in reality, there is no perfect experiment. So how do we know if the data are of sufficient quality? The answer often lies with the person who will ultimately use those data. How much do they need to trust the result in order to make the correct decision? Therein lies the need to provide not only the data, but a reasonable estimate of its uncertainty.

Normally, this information is presented as an interval inside of which we expect the true answer to lie along with a probability of this statement being correct. For example, we may say that the pressure is 1025 kPa  $\pm 20$  kPa with 95% confidence. A decision maker can then assess whether the uncertainty (20 kPa in this example) is sufficiently small to permit

their decision. If not, additional resources will be required to reduce the uncertainty by using better instruments and/or acquiring more data.

Consider this relatable example. Many automobiles measure their remaining fuel and fuel burn rate and combine these measurements to report the current range of the automobile. A driver uses this information to decide when to refuel. Say Steve is driving in rural New Mexico and his car reports a 40 mile range while he is passing the last gas station for the next 35 miles. If the range measurement is perfect, there is no reason to be concerned. But no one would assume this estimate is perfect. What if the car reported the range as  $40 \pm 5$  miles? This estimate seems to open the possibility that he would be pushing the car into the gas station.

Since that is an unpleasant possibility, Steve will naturally wonder how confident we are in the estimate of  $\pm 5$  miles. If Steve was told that we were 68% sure of the estimate, he may elect to refuel the car now. If Steve insisted on a better estimate, perhaps  $\pm 10$  miles could be reported as an estimate we are 95% sure of. Then Steve may realize that to be 95% sure, the “cushion” on the fuel is no longer acceptable. This is how uncertainty is used by decision makers.

Uncertainty is also a tool used to decide if measured results are believable. This point can be illustrated by a pragmatic example that comes from the author’s own experience. Figure 2.2 is a simplified version of the results from the study of minor losses in a bank of tubes [4]. The minor loss coefficient is defined as

$$K = \frac{2\Delta P}{\rho U^2}. \quad (2.1)$$

where  $\rho$  is the fluid density,  $\Delta P$  is the change in pressure from one tube to the next, and  $U$  is the maximum velocity between the tubes. This quantity is shown as a function of Reynolds number  $Re = UD\rho/\mu$ , where  $D$  is the diameter of the tubes and  $\mu$  is the fluid viscosity. So to find the loss factor, the material property  $\rho$  must be determined and the pressure drop and velocity must be measured.

It is easy to see in Fig. 2.2 that these measurements generally resembled previous results found in the literature (as shown by the blue line). However, as the Reynolds number became small, the results suddenly departed from that behavior. Was the sharp upward trend seen on the left-hand side of the plot interesting physics or indicative of a problem with the

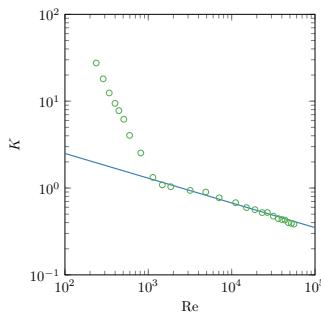


Figure 2.2: Minor loss data from [4]. The symbols are measurements from that study while the line is a representation of other results in the literature.

measurements? One could come to a variety of conclusions. Perhaps the flow at smaller Reynolds numbers is laminar and becomes turbulent at  $\text{Re} = 1000$ .<sup>\*</sup> Questions would surely arise since the results of previous studies, as shown by the solid line, indicate a different trend. Should we question our measurements or the previous work?

Carrying out an *uncertainty analysis*, using the tools described in the following chapters, provides some important insight into these data, as shown in Fig. 2.3. These are 95% confidence uncertainty bands, so one expects that 95 times out of 100, the unknown, true, answer lies within these bands.

When the data are presented with *uncertainty bands*, the answer to the question of whether the trend is interesting physics or measurement issues becomes apparent, as the uncertainty bands are considerably larger than the presented values for  $K$ . In this specific case, the issue was the pressure measurement had become so small that it could not longer be resolved by the pressure transducer. The pressure data below a certain value appeared to be constant since it was below the noise floor of the device! At the same time, the velocity continued to decrease so a distinct power-law trend, (which was *not* physical) appeared in the plotted results. In addition to informing any one who may wish to use these data on the expected quality,

---

\*BLS formulated a physics-based explanation of these trend prior to performing the uncertainty analysis and is happy that no physical evidence of this argument exists today.

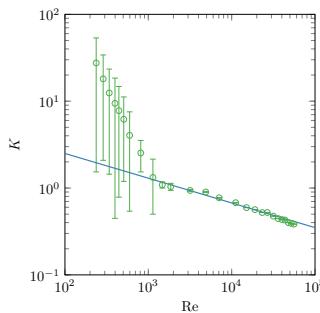


Figure 2.3: Uncertainty bands on the data help to give perspective on the role of physics versus measurement uncertainty.

it also provides critical information for the experimentalists, who can then repeat the experiment by using a pressure transducer with a different range or possibly make use of two or more different pressure transducers.

## 2.1 Errors

Experimental errors cannot be eliminated. Furthermore, reducing them usually comes at a cost. One key purpose of uncertainty analysis is to guarantee that resources are spent on the most important error sources and to ensure that the overall uncertainty of the experiment is within acceptable limits.

It is customary to divide error sources into at least two categories and many such divisions exist in the literature. In each case, the motivation for this division is to separate error sources by their effect on the final result or how they are treated.

The most common categories used in the USA are *bias* (or *systematic*) and *random* errors. Different definitions of these may also be found. According to Coleman and Steele [2], the traditional U.S. categorization of random and systematic errors states that a systematic error does not change its value during the measurement period while a random error does. In the 3rd editions of their text, Coleman and Steele had used this definition [1]. However, as they note in the 4th edition [2], this definition presents the difficulty that random errors defined this way may not actually vary randomly (e.g. experimental drift meets this definition of random

error).

The International Organization for Standardization produced the Guide to the Expression of Uncertainty in Measurement (GUM) in 1995 [5]. This was likely the most ambitious attempt to organize and standardize the expression of uncertainty. This document has since been updated (with minor corrections) in 2008 by the Joint Committee for Guides in Metrology (JCGM), which is comprised of members from several organizations, including the ISO [6]. The JCGM also produced several companion documents, including the 2009 document entitled *Evaluation of Measurement Data — An Introduction to the “Guide to the Expression of Uncertainty in Measurement”*. This document states: “There are two types of measurement error quantity, systematic and random. A systematic error (an estimate of which is known as a measurement bias) is associated with the fact that a measured quantity value contains an offset. A random error is associated with the fact that when a measurement is repeated, it will generally provide a measured quantity value that is different from the previous value. It is random in that the next measured quantity value cannot be predicted exactly from previous such values. (If a prediction were possible, allowance for the effect could be made!)” [7]. As we will demonstrate below, there is a continuous spectrum of error behavior from randomly varying during an experiment to remaining fixed.

*Systematic* has also been used to describe an error that does not change its value with fixed, known inputs. For example, a sensor may have a sensitivity that is a function of ambient temperature. Meanwhile, ambient temperature may vary due to the cycle of the HVAC system. In this example, the bias changes its value during the measurement period. However, as its relationship to a known input, such as room temperature, is known, it may be considered systematic. From this point of view, biases can become random if the relevant inputs are known and randomized. It should be noted that there may be varying error sources that cannot be randomized.

In many cases, we are actually discussing errors that vary at different rates. These differences may require different methods to quantify their effect (which is to say to assess the uncertainty), but categorizing them as a different kind of effect has little utility.

We will explain these concepts through line plots of a measurement over several samples, as shown in Fig. 2.4. The abscissa ( $x$ -axis) in each

of these is the sample number, and these samples could be acquired over time or space at equal intervals or not. In each case, the measurement suffers from two error sources,  $\delta_r$  and  $\delta_s$ . While this text will not label errors as systematic or random,  $\delta_s$  is an error source that would typically be called systematic, since, in most cases, it varies slowly or not at all. The  $\delta_r$  source varies randomly in every sample. The various plots show examples of measurements with different magnitudes and behaviors of each error source.

The effect of errors that vary randomly and are acquired many times may be mitigated through averaging. Therefore, if a mean value is sufficient, and sufficient data points are available, then randomly varying errors may be of little concern. In this example, the effect of  $\delta_r$  can be mitigated through averaging.

This is not true of an error that is fixed, which cannot be detected in the measurement. It is also important to note that error sources that have traditionally been categorized as “bias errors” can vary. The variation may be random or not. In Fig. 2.5, the  $\delta_s$  varies during the measurement either systematically or randomly. This concept can be difficult to understand, but it is easy to find examples of each of these phenomena. A systematically varying bias may be generated by instruments that have an error that is a function of room temperature. If the room temperature follows a typical HVAC behavior (a sawtooth wave), the bias error will be correlated to the temperature, as shown in Fig. 2.5a.

Errors are, in general, random phenomena. Decades of previous literature on measurement uncertainty have taught us to believe that some errors do not vary at all. In reality, so-called “bias errors” are fixed because they may be observed only a single time. Using an example from [2], if one makes a voltage measurement with a meter, there is an error associated with the factory calibration of that particular unit. If one were to acquire another meter of the same model that was calibrated separately, one would acquire a new sample of this error source and the error would be different. If this process is repeated sufficiently, the random nature of the calibration error would become apparent. The standard deviation of the “bias errors” of all the units should scale with the advertised accuracy of the meter. Coleman and Steele call this phenomenon “fossilized” systematic error. In most cases, errors that have been termed bias errors are functions of known and/or controlled parameters, or perhaps the calibration source,

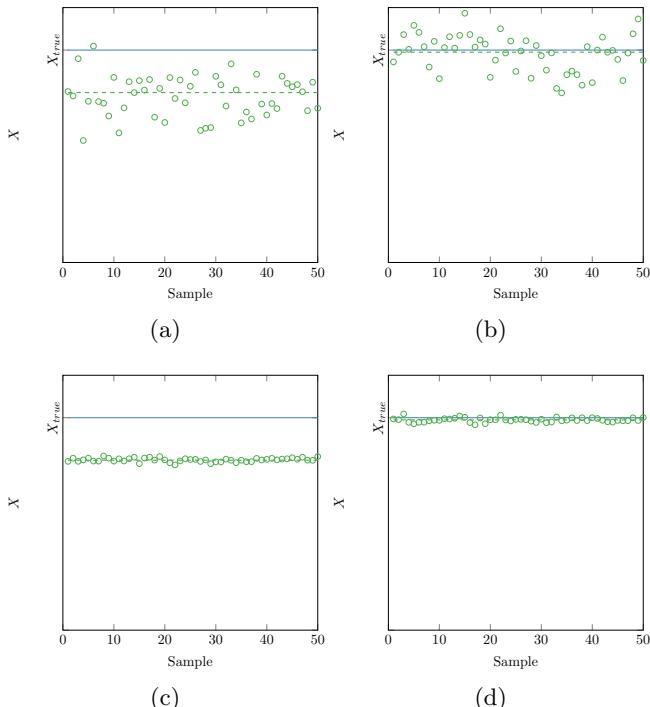


Figure 2.4: Examples of measurements with two error sources, one of which is resampled during the measurement period and one which is not. The true value is the blue horizontal line. The fixed  $\delta_s$  error added to the true value is the green dashed line. a) A large  $\delta_s$  and large  $\delta_r$  error, b) A small  $\delta_s$  error and large  $\delta_r$  error, c) a large  $\delta_s$  error and small  $\delta_r$  error and d) a small  $\delta_s$  error and small  $\delta_r$  error.

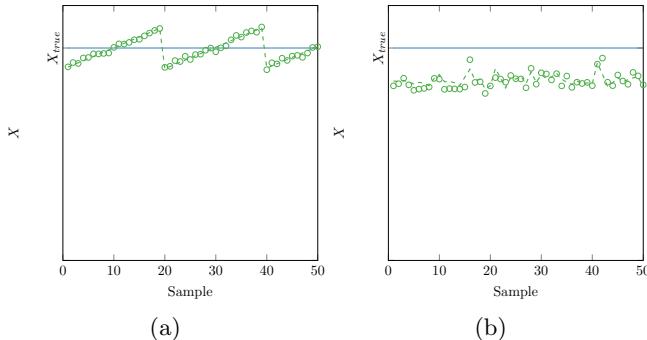


Figure 2.5: A measurement with a varying  $\delta_s$  error that changes its value during the measurement. The true value is the blue horizontal line. The  $\delta_s$  error added to the true value is the green dashed line. a) a systematically varying  $\delta_s$  error and b) a randomly varying  $\delta_s$  error.

while so-called random errors stem from uncontrolled or unknown sources.

Although less common, it is possible to have an error source that depends on a known parameter that varies randomly. The fact that this source is a predictable function of a known input would put it into the category of a bias error in many nomenclature schemes [8], but we believe there is no benefit to this categorization since all that matters is the manner in which this error source is analyzed.

Additionally,  $X_{true}$  can vary with sample number (e.g. time or space). This variation could be random, systematic, or periodic. An example is shown in Fig. 2.6. In this case, the true value of  $X$  varies in a periodic fashion while the  $\delta_s$  is constant. In Chapter 4, we will discuss methods to find the *phase-averaged* value, corresponding to the red dashed line. Of course, there are many ways that  $X_{true}$  can vary (randomly, systematically, periodically with a period longer than the measurement, etc.).

It is important to be mindful of the fact that these are all constructs. As stated by Coleman and Steele, “Errors do not know what their names are!” (pg. 14, 4th ed.). We believe it is wrongheaded to describe any of these schemes as incorrect. We do claim, however, that some are more useful than others. We choose to treat all errors as random and delineate systematic and random uncertainties according to whether they are based upon sampling or not.

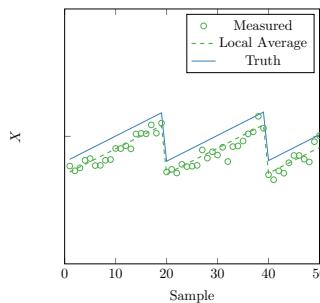


Figure 2.6: A measurement of a process that varies periodically with sample number. The measurement is impacted by a small bias (which shifts the measurements downward) as well as random error.

## 2.2 Uncertainty

In this text, we will acknowledge the random nature of most errors from the start. As such, we will not distinguish between bias (systematic) and random error sources or uncertainties. The uncertainty of any quantity will be found as the root sum of all of the quantity's uncertainty sources. One of these may or may not derive from the statistics of a sample. For a measured quantity  $X$  with  $n$  known error sources,

$$u_X = \sqrt{r_X^2 + u_{X_1}^2 + u_{X_2}^2 + \dots u_{X_n}^2}, \quad (2.2)$$

where  $r_X$  is the random uncertainty, possibly based on sample statistics, and  $u_{X_i}$  is the uncertainty due to the  $i$ th error source of  $X$ . Methods for finding  $r_X$  are discussed in Sec. 4.1 while the  $u_{X_i}$  values are discussed in Sec. 4.2.

The modeling uncertainty community uses the categories of *epistemic* and *aleatory* uncertainties. Epistemic uncertainties are defined as those stemming from a lack of knowledge while aleatory uncertainties are due to probabilistic variations. Clearly, these categories are similar to bias and random uncertainties, although there are limits to that analogy. Perhaps the most important drawback of these categories is the fact that the terms are not descriptive to most readers and while certainly not intentional, they drive an unnecessary wedge between the measurement and modeling community.

Moffat [9] discussed the “replication” of an experiment and the use of three different hypothetical “levels” of replication. While this paradigm is rooted in the notion of steady experiments, we bring it up here since it 1) is used extensively in [1] and 2) aids with the interpretation of a Monte Carlo Method (MCM) uncertainty analysis presented in Chapter 5. First, an  $N$ th-order analysis imagines that for each sample, all facilities and instruments are replaced with one of the same specifications. As a result, all bias errors would be resampled in addition to random errors. For a 1st order analysis, only time, space, or sample number (i.e. the independent variable of the measurement) is allowed to vary while all biases remain fixed. For a zeroth order analysis, only the resolution of the instruments is considered.

Coleman and Steele [1] use this concept to separate fluctuations in  $X_{\text{true}}$  from fluctuations in the measurement due to errors. Since 1st level uncertainty is simply the standard deviation of the experimental result, they claim that the  $N$ th order uncertainty minus the first order uncertainty reveals the fluctuations in  $X_{\text{true}}$ . In the present text, we prefer to acknowledge that  $X_{\text{true}}$  can vary and build nomenclature to accommodate that possibility. However, we will use Moffat’s ideas to help interpret MCM results.

## 2.3 Nomenclature Used in This Text

We define an error  $\delta$  to be the difference between a measurement and the unknown true value being measured  $\delta_X = X - X_{\text{true}}$ . While we will not attempt to segregate stationary error sources from varying ones, the total error may be averaged, and this average may be removed to provide a zero-mean error  $\epsilon_X = \delta_X - \bar{\delta}_X$ . This will be useful when defining covariances in Chapter 5.

Uncertainty is an estimate of the width of distribution of errors. In most cases, we assess that width through the standard deviation  $s_{\delta_X}$ . We note that the standard deviation operator is defined in Chapter 3 in Eq. 3.14.

Most texts on engineering measurements do not emphasize the measurement of processes that vary in time. In [2], there is discussion of experiments where the true value of the quantity being measured changes from one sample to the next, which they call a “Sample-to-Sample Experiment.” This text will treat such experiments in the same way as processes

that vary in time.

$$X(t) = X_{\text{true}}(t) + \delta_X(t). \quad (2.3)$$

Therefore, if the standard deviation of the measured quantity is computed, we obtain the standard deviation of the true value plus the error. We note that both quantities can play an equal role. If  $X_{\text{true}}$  is a constant, then the standard deviation of the sample is equal to the standard deviation of the errors,  $s_X = s_{\delta_X}$ . The variation could alternatively be in space or from one sample to the next.

We will use  $u_X$ , to indicate uncertainty that is not derived from statistics. Total uncertainty will be denoted as  $u_X$  or  $U_X$  depending on the confidence level, as described in Chapter 4.

Any uncertainty, whether it be random, systematic, or total, is expressed as a band around the nominal value, or as  $X \pm u_X$ . This was termed in [1, 2] as an *uncertainty band*, counter to the more frequently used term “error bar.”<sup>\*</sup> We agree that error bar is a problematic term (although ubiquitous!) since it implies that it is a bound beyond which an error does not lie. This is the opposite of the truth, since for any confidence level used, errors always lie beyond the uncertainty band.

For this reason, uncertainty analysis is typically based on a “confidence level”, which says that some majority percentage of errors are smaller than uncertainty bands. There are many choices for confidence levels, but for various reasons discussed in Chapters 3 and 4, most of the time in engineering and physics, a confidence of 68% or 95% is used.

All plotting software we have encountered uses the term “error bar.” It is common for error bars to be used to indicate various quantities on plots. For instance, in some fields, it is common to use an error bar to express  $s_X$ . In perhaps rarer cases, they express a known bound on the error. It is always critical to be clear on the meaning of error bars on plots, including the confidence level used in a statistical uncertainty analysis.

These are illustrated in Figs. [2.7] a-b where a true signal (blue) is measured 100 times and subjected to noise. The measured data points are green. In Fig. [2.7] a, uncertainty bands with a 68% confidence level

---

<sup>\*</sup>In his workshop with co-author Glenn Steele, Hugh Coleman was known to remark “They are not error bars. They are uncertainty bands. Error bars are where errors go to drink beer and listen to music.”

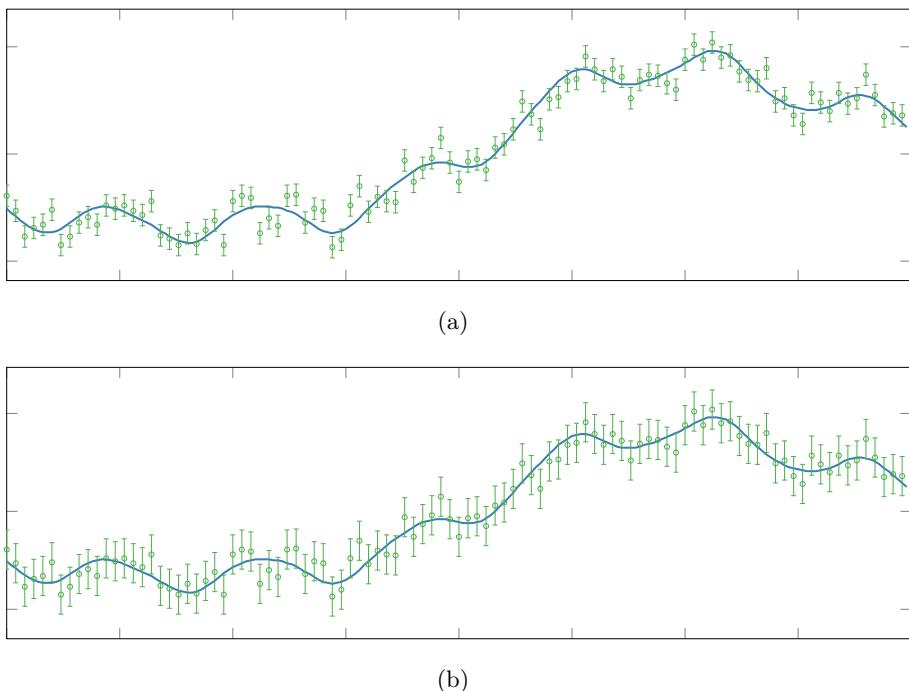


Figure 2.7: A true signal (blue) and 100 measurements of the signal subjected to errors (green symbols). a) 68% confidence level uncertainty bands and b) 95% confidence level uncertainty bands.

are shown while in Fig 2.7 b the confidence level is increased to 95%. We expect that all but 32 data points will have uncertainty bands that cover the actual signal for a 68% confidence interval and all but 5 for a 95% confidence interval. Note that in neither case is the truth always contained within the uncertainty bands. If we desire more of the uncertainty bands to cover the true signal, we could increase the confidence level, and the uncertainty bands would become larger.

The plotting style used in this text will attempt to guide the reader by using colors to indicate the common notions of error, truth, and measurements subjected to error. An example is shown in Fig 2.8. Generally, errors will be shown in red. True values will be indicated in blue, and measurements of the truth subjected to errors will be in green.

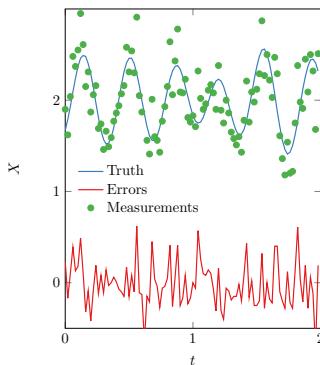


Figure 2.8: An example plot used to demonstrate the meaning of colors of the plots in this text. A true value (blue) is contaminated with errors (red) in the measurement process resulting in the green measured points.

## 2.4 Propagation Techniques

We will primarily use two tools to determine uncertainties in values that are computed based on measurements of more than one quantity, whether those multiple quantities consist of different measurements or elements of a mathematical operator.

### 2.4.1 Taylor Series Method

It is intuitive that any quantity formed by multiple other quantities is impacted by changes in those quantities. Clearly, the size of this impact depends on the sensitivity to the input quantities and the size of the perturbations to each quantity.

Taylor Series Method (TSM) can be used either for the propagation of multiple measured quantities (demonstrated in Chapter 5) or to determine the impact of an operator. TSM will be used in Chapter 4 to show how each measurement impacts the uncertainty of the mean. The method will be presented several ways in this text, but as an introduction, we will write it down for the uncertainty of a quantity  $Y$  that is a function of an arbitrary number of quantities  $X_i$ . The relationship between  $Y$  and the  $X_i$  values is called the Data Reduction Equation (DRE).

$$u_Y^2 = \sum_{i=1}^j \theta_{X_i}^2 u_{X_i}^2 + 2 \sum_{i=1}^{j-1} \sum_{k=i+1}^j \theta_{X_i} \theta_{X_k} u_{X_i X_k} \quad (2.4)$$

The  $\theta$  terms are the sensitivities of  $Y$  to each  $X_i$ , or  $\theta_{X_i} = \partial Y / \partial X_i$ . The second term involves correlations between the errors of the various  $X_i$  values ( $u_{X_i X_j}$ ), and may be ignored in some situations. This will be discussed in detail in Chapter 5.

The manner in which Eq. 2.4 is used will now be illustrated in two examples. Most commonly, we wish to know the uncertainty of a quantity that depends on other measured quantities. For instance, say that an experimentalist desires to determine the density of an ideal gas in a container. The temperature  $T$  and pressure  $P$  inside the container can be measured and the ideal gas law can be used to relate these to the gas density  $\rho$ :

$$\rho = \frac{P}{RT} \quad (2.5)$$

where  $R$  is the gas constant which is known exactly. Applying Eq. 2.4 to this relationship (ignoring the correlated terms), we find that

$$\begin{aligned} u_\rho^2 &= \left( \frac{\partial \rho}{\partial P} \right)^2 (u_p)^2 + \left( \frac{\partial \rho}{\partial T} \right)^2 (u_T)^2 + \left( \frac{\partial \rho}{\partial R} \right)^2 (u_R)^2 \\ &= \left( \frac{u_p}{RT} \right)^2 + \left( -\frac{u_T P}{RT^2} \right)^2 \end{aligned} \quad (2.6)$$

The negative in the second term will have no effect since that term is squared. Methods for finding the uncertainties for  $\rho$  and  $P$  are the subject of Chapter 4 while more details on propagation will be presented in Chapter 5.

As a second example, consider applying Eq. 2.4 to the mean operator. For  $N$  values of  $X_i$ ,

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i. \quad (2.7)$$

Assuming no correlation between the  $X_i$  values, the uncertainty of  $\bar{X}$  can be found by applying Eq. 2.4 to the mean operator.

$$u_{\bar{X}}^2 = \left(\frac{u_{X_1}}{N}\right)^2 + \left(\frac{u_{X_2}}{N}\right)^2 + \dots + \left(\frac{u_{X_N}}{N}\right)^2 = \frac{1}{N^2} \sum_{i=1}^N u_{X_i}^2. \quad (2.8)$$

For a case where the uncertainties of all values of  $X_i$  are the same and equal to  $u_X$ ,

$$u_{\bar{X}} = \frac{u_X}{\sqrt{N}}, \quad (2.9)$$

which says that the uncertainty of the mean depends on the uncertainty of the data points and decreases like the square root of  $N$ . This idea will be discussed extensively in Sec. 4.1.1.

## 2.4.2 Monte Carlo Method

In many cases, it is easier to analyze the impact of multiple error sources numerically instead of the analytical approach of TSM. The primary technique for numerical analysis of uncertainty is called the Monte Carlo Method (MCM), and its usefulness is broader than just uncertainty analysis. The MCM can be used for any relationship between variables even if some of them have a non-deterministic (i.e. random) behavior. This method does not require that the variables all have similar random behavior and also does not require the derivatives of the governing relationship to be determined.

In MCM, thousands of iterations are performed. For each iteration, random values of all of the variables in DRE are generated. Consider a simple example where  $z = f(a, b, c)$ . If the input variables are not known exactly, we can simulate the impact of their possible outcomes iteratively using MCM. As a simple way of visualizing a random number generator, consider the roulette wheels shown in Fig. 2.9.

There is a unique wheel for each variable. Looking at the  $a$  wheel, it is clear that  $1 \leq a \leq 18$  and that each of these values are equally likely. The variable  $b$  has the range  $-5 \leq b \leq 6$ , but the values near 0 are more likely to occur than the extreme positive and negative values. On the other hand,  $1 \leq c \leq 16$  and the possibility of each of these values is arbitrary. A major feature of MCM is the fact that although these three variables are distributed differently, their effect can be easily combined. The TSM is not capable of propagating variables with different distributions.

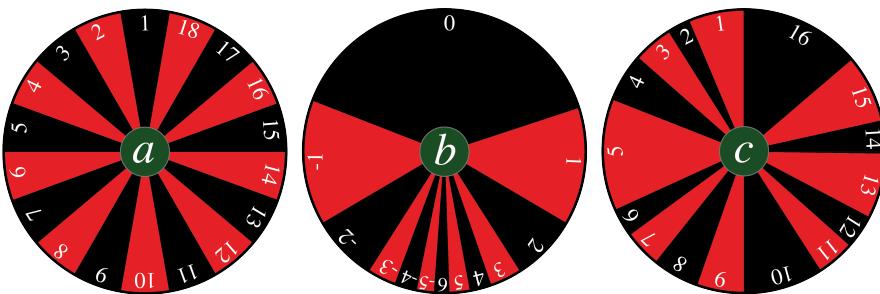


Figure 2.9: Three roulette wheels, one for each variable *a*, *b*, and *c*. The variable *a* has a uniform distribution while *b* has a central tendency (0 is much more likely than  $\pm 5$ ). For *c*, each value has a unique probability.

To perform an MCM, the three wheels are spun to generate a random value of  $a_1, b_1$  and  $c_1$  that can be used to compute  $z_1$ . A second spin can be used to compute  $z_2$ , and so on. The purpose of the analysis is to examine the properties of the resultant values of  $z$ . In using MCM for measurement uncertainty, we normally interpret *a*, *b* and *c* as simulated measurements that are contaminated with error and the width of the distribution of  $z$  as a metric of its resulting uncertainty. We commonly use a random number generator rather than a wheel, and the type of random number (e.g normal, uniform, etc., as described in Chapter 3) is based on our knowledge of the errors to which a given variable is subject. Monte Carlo simulations will be used to demonstrate the Central Limit Theorem in Chapter 3, and to propagate multiple measurements into a single, total uncertainty in Chapter 5.



# Chapter 3

# Probability and Statistics

Many texts on measurement uncertainty provide a more complete treatment of statistics than will be presented here. We choose to emphasize the aspects of statistics that are most useful to measurement uncertainty.

In order to assess the impact of errors on uncertainty, it is first necessary to know how they are distributed. This is commonly done with the aid of the *Probability Distribution Function*, or PDF. The familiar Gaussian distribution is a common example of a PDF. In what follows, we will outline how to derive a PDF for a process.

## 3.1 Characterization of Random Variables

Consider a random variable  $X$ . The probability that  $X$  takes on a value less than or equal to  $x$  is given by

$$p = P(X \leq x), \quad (3.1)$$

where  $p$  is a real number ( $0 \leq p \leq 1$ ) representing the likelihood that  $X \leq x$  occurs. An impossible condition results in zero and a definite condition results in unity.

$$P(X \leq x) = F_X(x). \quad (3.2)$$

The probability that the random variable  $X$  takes on a value less than or equal to  $x$  is determined by its distribution function,  $F_X(x)$ , which is

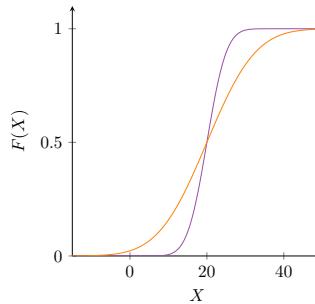


Figure 3.1: Cumulative Distribution Functions (CDF) for a Normal Distribution

called the *cumulative distribution function* (CDF) of  $X$ . The probability that  $X$  lies within the interval  $(a, b]$  is

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned} \quad (3.3)$$

In Eq. 3.3, the use of  $<$  and  $\leq$  on the left-hand side is important since this could make a difference in the case of discrete random variables. This point is discussed in more detail in many statistics books, including [10]. It is not of critical importance for understanding distributions of a random variable, but it is pointed out for completeness. Some basic properties of any CDF,  $F_X$ , are that it is non-decreasing and right continuous. The non-decreasing property has the definition:

$$F_X(b) \geq F_X(a) \text{ for } b > a \quad (3.4)$$

This property of the CDF is important in characterizing a plot of the CDF. As shown in Fig. 3.1, as  $x$  is increased, the CDF cannot decrease, but rather increase or stay the same. A CDF also has the properties:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad (3.5)$$

$$\lim_{x \rightarrow \infty} F_X(x) = 1 \quad (3.6)$$

Every function with these four properties is a CDF. For such a function, a random variable can be defined such that the function is cumulative distribution function of that random variable. Also, since the probability of every event is non-negative, i.e.

$$F_X(b) - F_X(a) = P(a < X \leq b) \geq 0 \quad (3.7)$$

The *probability density function* (PDF) of a continuous random variable can be determined from the CDF by differentiating using the Fundamental Theorem of Calculus; i.e. given  $F_X(x)$ ,

$$f(x) = \frac{dF_X(x)}{dx}. \quad (3.8)$$

A property of the PDF is that it satisfies the normalization condition. So, for any PDF,

$$\int_{-\infty}^{\infty} f(x)dx = 1. \quad (3.9)$$

This condition means the total area under the curve must not be greater than 1. Combining the above definitions of probability, CDF and PDF, the probability that random variable  $X$ , assumes a value that lies in the interval  $(a, b]$ .

$$P(a < X \leq b) = \int_a^b f(x)dx. \quad (3.10)$$

A general example is shown in Fig. 3.2. The integral of a PDF over some interval of  $X$  is the probability of a value of  $X$  within that interval.

Random events are often described through probability density functions and their parameters, which are often statistical quantities. The most important statistics to measurement uncertainty are the mean and standard deviation (or square root of the variance). A random variable  $X$  may have a *parent mean*, denoted by  $\mu$ , and *parent standard deviation*, denoted by  $\sigma$ , which are defined as

$$\mu_X = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i \quad (3.11)$$

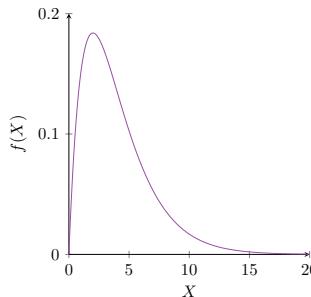


Figure 3.2: A generic probability density function (PDF)

and

$$\sigma_X = \lim_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 \right]^{1/2} \quad (3.12)$$

respectively. These parent statistics would be the result of having access to an infinite number of samples which in practice is not possible. It is common to use Greek symbols to represent these quantities that cannot be known exactly. We can estimate  $\mu$  and  $\sigma$  based on samples,

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i, \quad (3.13)$$

and

$$s_X = \left[ \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \right]^{1/2}, \quad (3.14)$$

where  $s_X$  is the *sample standard deviation* and  $\bar{X}$  is the *sample mean*. Both of these arise from a finite number of measurements acquired from the parent population. This subset is referred to as a *sampling distribution*. It is customary to use the Roman variables for these estimates of the parent distribution quantities.

The mean and standard deviation are parameters of some PDFs, and for some distributions, such as the Gaussian (or normal) distribution described below, may be sufficient to fully describe the PDF. Other, more complicated distributions, may require more parameters. We will begin

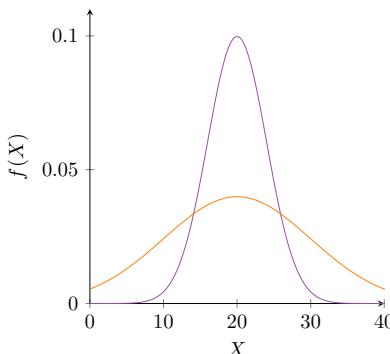


Figure 3.3: Shape of a Gaussian (normal) distribution. The mean  $\mu = 20$  and for the orange curve,  $\sigma = 10$  while the purple curve has  $\sigma = 4$ .

with a Gaussian PDF and expand to more sophisticated models as required by later topics.

## 3.2 The Gaussian Distribution

Most important random phenomena associated with measurements have a *central tendency*. Events far from that central tendency are less likely than events closer to it. In many important cases, events higher than the central tendency are as likely as events below it. One common model for a process with these features is the Gaussian, or normal, distribution.

For the Gaussian PDF,

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}, \quad (3.15)$$

where the parent statistics are the standard deviation  $\sigma$  and the mean  $\mu$ . The familiar “bell curve” shape of  $f$  can be seen in Figure 3.3. Note that since the inverse of the standard deviation multiplies the exponential and appears in the argument of the exponential, larger  $\sigma$  means  $f$  is wider and more squat. If we define  $\tau = (X - \mu)/\sigma$  as the number of standard deviations away from the mean, the probability of an event occurring is

$$P(\tau) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\tau}^{\tau} e^{-\tau^2/2} d\tau.$$

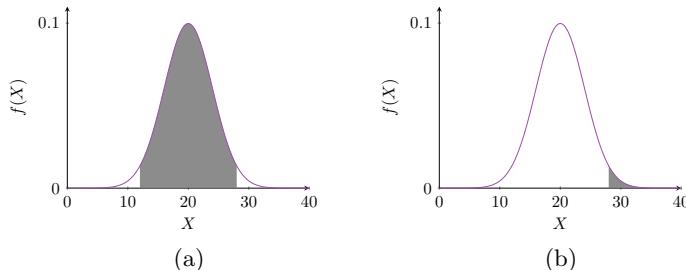


Figure 3.4: Gaussian distribution with mean  $\mu = 20$  and  $\sigma = 4$ . The shaded area is a) 12 to 28 representing the table lookup and b) 28 to  $\infty$  which is the value required in the example.

It is common to tabulate the integral of  $f$  up to various values of  $\tau$ . This integral represents the probability of an event occurring in the range  $-X < \tau < X$ . Such a table is provided in Appendix A.

For example, consider a Gaussian process that has a mean of 20 and a standard deviation of 4. If we wish to know the probability of this process generating a value of 28 or larger, we first find  $\tau = (28 - 20)/4 = 2$ . Turning to the Gaussian PDF table contained in Appendix A, we see that the probability is 95.45%, but note that this table is 2-sided. It provides the integral from  $-\tau$  to  $\tau$  as shown in Fig. 3.4a. In order to find the probability in the upper tail shown in Fig. 3.4b, we note that the Gaussian distribution is symmetric and the integral of any PDF from  $-\infty$  to  $\infty$  is 1, so  $P = (1 - 0.9545)/2 = 0.0228$  or 2.3%.

When considering measurement uncertainty, our interest in the normal distribution can be narrowed to knowledge of the values of  $\tau$  that correspond to confidence levels of interest. By confidence level, we mean the probability of an event happening that is a given number of standard deviations from the mean. In engineering applications, two common values are used: 68% which is the probability of an event happening one standard deviation from the mean ( $\tau = 1$ ) and 95% , for which  $\tau = 1.96$ . These are frequently called  $1\sigma$  and  $2\sigma$  values.

Note that the Gaussian PDF allows for the possibility of events that are very far from the mean. This is the reason that expressing uncertainty as an interval inside of which we are certain the true value lies is not useful.

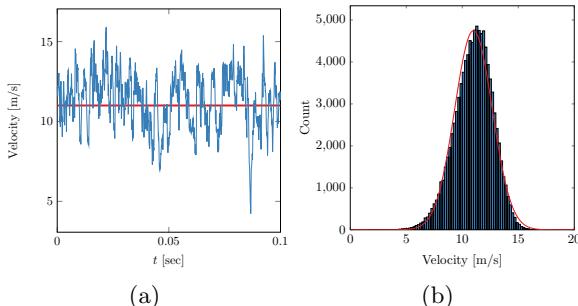


Figure 3.5: A small sample of velocity as a function of time in a turbulent jet. Histogram of 100,000 samples in the same jet.

If we assume a Gaussian PDF, such bands would need to be infinitely wide.

Strictly speaking, few real processes follow a Gaussian PDF. For instance, for a process to be strictly Gaussian it must be possible for any value to occur in either direction, above or below the mean.

Consider turbulent fluid flow. The velocity magnitude at a point in a turbulent flow may appear very Gaussian, but for it to be Gaussian, non-physical velocities must be possible. Generally, it is much more likely for a process to follow a Gaussian distribution for events that occur between the mean and one standard deviation from the mean than events that are less probable.

Figure 3.5 is a histogram of turbulent jet velocity data from [11]. The record of 100,000 samples is divided into 100 bins, resulting in a distribution that appears continuous and Gaussian. For reference, a Gaussian distribution with the same statistics ( $\bar{V} = 11.0$  m/s,  $s_V = 1.68$  m/s) is superimposed. The velocity in this jet is more likely to be somewhat higher than the mean than somewhat lower. This is evidenced by the gap between the histogram and the Gaussian distribution just below the mean. The velocity is also more likely to be much lower than the mean than a similar amount higher than the mean as evidenced by a similar gap near 15 m/s. The actual distribution is not symmetric like the Gaussian. in spite of these differences, for many applications, it would be useful treat this velocity as a Gaussian process.

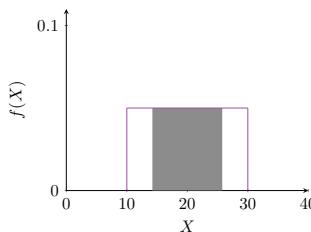


Figure 3.6: Uniform distribution from 10 to 30. The shaded area is  $\pm 1\sigma$ .

### 3.3 Uniform Distribution

There are many distributions that errors could follow other than a Gaussian distribution. The most prevalent distribution in measurements other than the Gaussian distribution is a uniform distribution, meaning that any value is as likely as any other within some bound. This is pictured in Fig. 3.6. Since a PDF must integrate to unity, a uniform PDF of width  $2A$  has a height of  $1/2A$ . In this case,  $\mu = 20$  and  $A = 10$ . The standard deviation of a uniformly distributed variable of width  $2A$  is  $A/\sqrt{3}$ . Digitizing errors, which are common but usually small, have a uniform distribution.

### 3.4 Distributions from Periodic Error Sources

As will be discussed in Sec. 4.2.2, many error sources are sensitive to environmental factors (e.g. temperature), and these often follow a cyclic pattern. For instance, indoor temperature may be a sawtooth wave with a period that is set by the heat load and insulation of the building. Outdoor temperature is generally somewhat sinusoidal. For this reason, we require the PDF that results from a cyclic function.

#### 3.4.1 Distributions from a Sinusoidal Error Source

For an error that takes on a sinusoidal nature, we will need to determine the PDF of a sinusoid. A sine wave that is defined as:

$$X = \sin(Y), \quad (3.16)$$

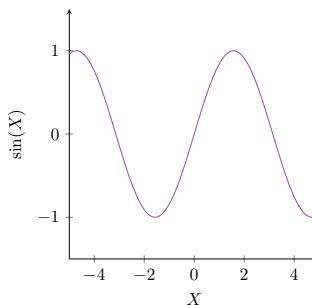


Figure 3.7: A sine wave as an example of a cyclic error source.

The sine function is shown in Fig. 3.7. From Eq. 3.16,  $Y$  is uniformly distributed over  $[-\pi/2, \pi/2]$ , so the resulting range of  $X$  values is  $[-1, 1]$ .

Here  $X$  is our random variable for which the probability density function is given by Eq. 3.10. In order to get the PDF, the function  $f(x)$  must be determined for Eq. 3.16. The inverse of a function provides the distribution of these values that can be cast into a CDF. Taking the random variable  $X$  which can have a value with the interval  $[-1, 1]$ :

$$P(X < a) = P(\sin(Y) < a) \quad (3.17)$$

$$= P(Y < \arcsin(a)) \text{ where } -1 \leq a \leq 1. \quad (3.18)$$

This equation is modified by adding a dividing factor of  $\pi$  and an additive factor or  $1/2$  which make it adhere to the properties of the CDF as described in Sec. 3.1. The above equation then becomes:

$$P(X < a) = F(a) = \frac{\arcsin(a) + (\pi/2)}{\pi}. \quad (3.19)$$

The CDF of a sine wave is shown in Fig. 3.8. The possible values that a sine function can have are shown along the abscissa of Fig. 3.8 while the probability is shown along the ordinate. As discussed in Sec. 3.1, the maximum probability is 1.

Finally, recalling Eq. 3.8, the PDF of a sine wave can calculated by taking the derivative of  $F(a)$  as shown in Eq. 3.19:

$$f(x) = \frac{d(F(a))}{dx} = \frac{1}{\pi\sqrt{1-a^2}}. \quad (3.20)$$

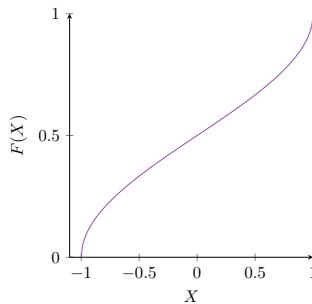


Figure 3.8: The Cumulative Distribution Function (CDF) of a sine wave.

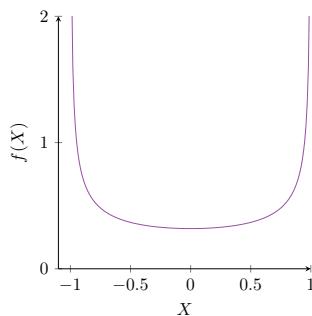


Figure 3.9: The Probability Density Function (PDF) of a sine wave.

A plot of the PDF for a sine wave,  $f(x)$ , is shown in Fig. 3.9. This plot shows a unique feature of the PDF of a sine wave. Specifically, as the slope of a function increases for a particular value, the probability of that value occurring increase significantly. Also note that this distribution is somewhat opposite of the central tendency of a Gaussian distribution. The extreme values are much more likely than the center.

In a random uncertainty analysis, the confidence interval for the standard uncertainty is the  $1 - \sigma$  value of the PDF of the error. Previously, the variance was introduced ( $\sigma^2$ ) as the second central moment of the distribution. In general if the PDF is known and can be integrated, then a formula for the variance exists (and subsequently, the square root of the variance – the standard deviation,  $\sigma$ ). The arcsine distribution is a special case of a family of distributions (2 parameters) called beta distributions. There is a formula for their variance in terms of their parameters:

$$\begin{aligned}\sigma^2 &= \frac{1}{8}(b-a)^2 && \text{for } -\infty \leq a \leq b \leq \infty \\ \sigma &= \frac{\sqrt{8}(b-a)}{8}.\end{aligned}\tag{3.21}$$

Equation 3.21 yields the confidence interval for the standard uncertainty  $(1-\sigma)$  over  $[a, b]$ .

### 3.4.2 Distributions From a Sawtooth Error Source

Another common type of error distribution is that of a triangle wave or sawtooth function. This was previously introduced in Sec. 2.1 and shown in Fig. 2.5. It is commonly found in errors that are caused by fluctuations in temperature and can be directly associated with the temperature variations from the ambient temperature in a room that is controlled by an HVAC system.

Deriving the PDF of a sawtooth function is similar to a sinusoid, but slightly more complicated since a sawtooth function must be handled in a piecewise manner:

$$y = f(x) = \begin{cases} \frac{y_o}{x_o}x & \text{for } 0 \leq x \leq x_o \\ \frac{y_o}{1-x_o}(x-1) & \text{for } x_o \leq x \leq 1 \end{cases}\tag{3.22}$$

Where each line segment of the sawtooth function is defined as a line segment over the range of  $x$ -values it occupies. Following the definition in Eq. 3.2, and if  $X$  is a random variable over the interval  $[0,1]$  that takes the form  $Y = f(x)$ , then:

$$F_Y(a) = P(Y \leq a) = P(f(X) \leq a)\tag{3.23}$$

The value  $a$  is any value of  $y_o$  where the probability and associated CDF and PDF are needed. This is shown graphically in Fig. 3.10.

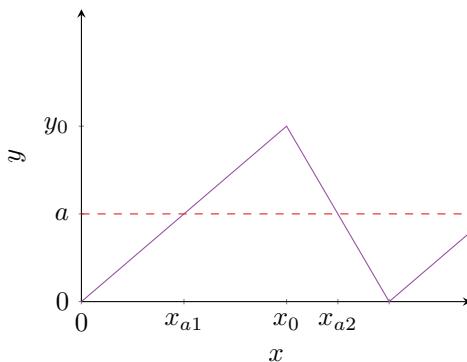


Figure 3.10: Probability of the values  $[0, a]$  existing on a sawtooth function

$$\begin{aligned}
 F_Y(a) &= x_{a,1} + 1 - x_{a,2} \quad \text{for } 0 \leq a \leq y_o \\
 &= \frac{ax_o}{y_o} + \frac{a(1-x_o)}{y_o} \\
 &= \frac{a}{y_o}
 \end{aligned} \tag{3.24}$$

Next, recalling the definition relating the CDF to the PDF (Eq. 3.8):

$$f_Y(a) = \frac{dF_Y(a)}{dx} = \frac{1}{y_o} \quad \text{for } 0 \leq a \leq y_o \tag{3.25}$$

This yields the interesting result that the PDF of a sawtooth function is a uniform distribution, as described in Sec. 3.3. The associated confidence interval  $(1 - \sigma)$  for the standard uncertainty can be determined using the properties of the uniform distribution.

A general approach for handling errors that vary periodically is to subtract out the mean value of the fluctuations. This approach accounts for the known error to be removed up front and then estimate the error caused by the fluctuating component. In this case, the range of the sawtooth wave is  $y_l \leq a \leq y_u$  and the corresponding CDF is:

$$F_Y(a) = \frac{a}{y_u - y_l} = \frac{1}{y_o} \quad \text{for } y_l \leq a \leq y_u \tag{3.26}$$

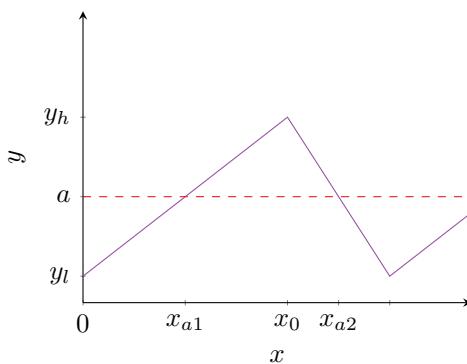


Figure 3.11: Sawtooth wave with the mean value subtracted out.

This can be differentiated using Eq. 3.8 to yield the PDF over the same range of values:

$$f_Y(a) = \frac{1}{y_u - y_l} = \frac{1}{y_o} \quad \text{for } y_l \leq a \leq y_u \quad (3.27)$$

## 3.5 Central Limit Theorem

Compared to most texts, the reader may find this text more oriented to Gaussian statistics. We make this choice for three reasons: 1) Teaching many PDFs to students who do not yet understand their purpose is confusing, 2) most errors follow a Gaussian distribution, at least up to  $1 - \sigma$  from the mean, and 3) even if they do not, their results can often be treated as Gaussian because of the Central Limit Theorem.

The Central Limit Theorem states that if a quantity  $Y$  is the sum of several random variables  $X_1, X_2 \dots X_n$ , then  $Y$  will be normally distributed as long as each  $X_i$  contributes similarly to the total, independent of the distributions of the input variables. Since most uncertainty quantities are sums of several other quantities, the assumptions of the Central Limit Theorem are often satisfied in measurement uncertainty.

This is easily shown by performing a Monte Carlo simulation (cf. 2.4.2) of  $Y = X_1 + X_2 + X_3 + X_4 + X_5$  by choosing random numbers between 0 and 1 from a uniform distribution for  $X_1$  to  $X_4$  and a triangular distribution for  $X_5$ . Histograms of each  $X$  and  $Y$  value are shown in Fig. 3.12. The

Gaussian nature of  $Y$  may seem surprising, but consider the fact that obtaining a small value of  $Y$  requires small values of all the  $X$  values, and is therefore unlikely. This is similarly true of large values of  $Y$ , while moderate values of  $Y$ , which can be formed by a variety of  $X$  values, are quite likely.

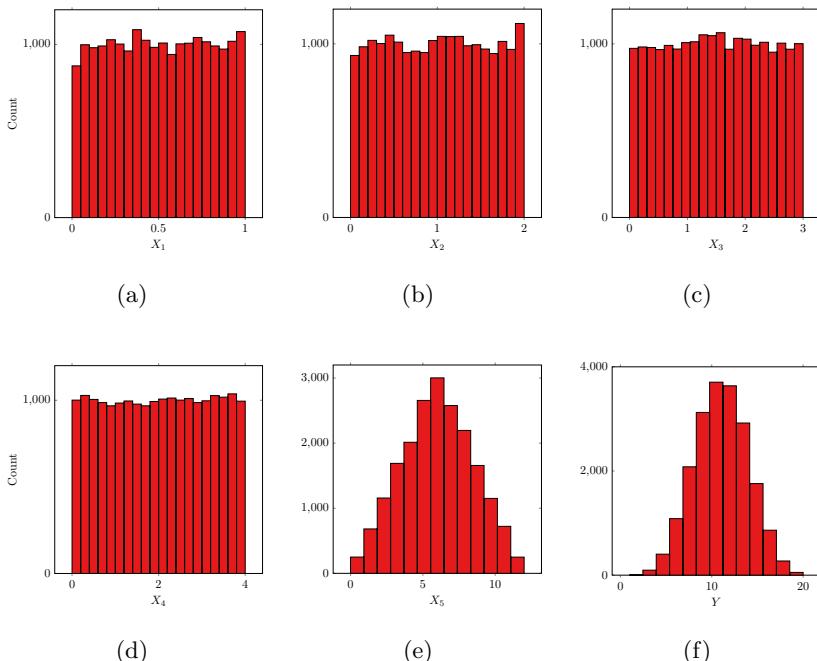


Figure 3.12: Histograms of random variables drawn from (a-d) an uniform distribution, (e) a triangular distribution and (f) their sum.

## Summary

You may test your understanding of this chapter by considering the answers to these questions:

- If a process has a Gaussian PDF with a mean of zero and a standard deviation of  $\sigma$ , which is the bound of possible results?

- Consider the possibility of a real process that is strictly Gaussian. What are the ramifications of this assumption?
- If a result consists of many measurements impacted by several error sources, why is Gaussian behavior a good assumption?
- What is an example of a PDF that does not have a central tendency?

## Homework Problems

- 1) If  $x_1, x_2, x_3, x_4$  and  $x_5$  are random variables with a uniform distribution of similar range, what is the distribution of  $x_1+x_2+x_3+x_4+x_5$ ? Why?
- 2) If  $X$  follows a Gaussian distribution with  $\mu = 2$  and  $\sigma = 7$ , what is the probability of a value above 9? Above 15? What is the upper limit on how large  $X$  can be?
- 3) Six dozen MLB baseball seam heights were measured resulting in a mean height of 33/1000 of an inch and a standard deviation of 3/1000 of an inch. By randomly choosing samples from a Gaussian distribution with those parameters, estimate the range of seam heights for 72 samples. What is the probability of a 20/1000 of an inch seam height?
- 4) You measure the diameter of 6 parts resulting in the tabulated value in mm. Compute the mean and the standard deviation.

4.1  
4.5  
3.9  
4.5  
3.8  
4.2

- 5) Visually estimate the standard deviation of the data shown in Fig. 3.13.

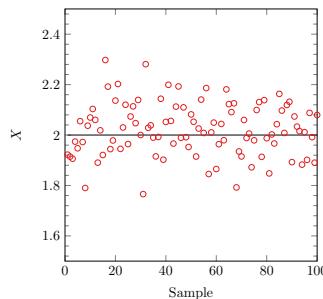


Figure 3.13

- 6) Consider the PDF of the sum of the roll of two dice. Sketch the PDF. In what ways is it similar or different than a Gaussian PDF?
- 7) Consider the distribution of any quantity you encounter everyday, such as the height of adults. Does this quantity follow a Gaussian PDF? Why or why not?
- 8) Estimate the PDF of a sinusoid by sampling uniform random values of  $x$  between 0 and 1 in  $X = \sin(2\pi x)$ . Plot the histogram of  $X$  and compare it to Fig. 3.9.
- 9) You plan to test the yield strength of a treated material. The test apparatus specifies an accuracy of 1% of reading and a non-repeatability of 1% of reading. You test 5 samples and obtain the following values.  
 300 MPa  
 280 MPa  
 310 MPa  
 250 MPa  
 255 MPa  
 What is the overall 95% confidence uncertainty of the average yield strength?

# Chapter 4

## Uncertainty of a Single Variable

By single variable, we refer to the measurement of a single measured quantity. This could be an instantaneous measurement of a process or the mean of a time-varying quantity. These single variables will form the inputs to the propagation of uncertainty discussed in Chapter 5.

Following Coleman and Steele [1] and others, we define a *standard uncertainty* of a single variable as a  $1 - \sigma$  (68% confidence for a Gaussian distribution) quantity. This uncertainty may be expanded to any desired confidence by multiplying by the appropriate coverage factor, as discussed in Chapter [3]. This requires an assumption about the distribution of errors, and the assumption of Gaussian behavior is commonly made; however, other distributions can exist and the most common types are discussed in Chapter [3].

Any quantity may be affected by multiple error sources. The resulting uncertainty due to each of these known sources must be determined and added to the total uncertainty of the measured quantity in a root-sum fashion. In addition, if multiple samples are acquired, the statistical uncertainty of each sample or the mean may be assessed. We will denote this quantity as the *random uncertainty* of  $X$ ,  $r_X$ . The uncertainties due to known error sources in addition to the statistical uncertainty will be combined to form a total uncertainty for a quantity. If  $X$  is affected by  $n$

known error sources and is sampled multiple times,

$$u_X^2 = r_X^2 + u_{X_1}^2 + u_{X_2}^2 + \dots + u_{X_n}^2. \quad (4.1)$$

It is often useful to note that any one term in a root sum that is less than 1/4 the size of any other will not have significant impact. This is referred to as the *1/4 rule*. In this chapter, we first discuss estimation of  $r_X^2$  followed by the  $u_{X_n}^2$ .

## 4.1 Random Uncertainty of a Single Variable

Recall from Eq. 2.3 that for a measurement of a time-varying process,  $X(t) = X_{\text{true}}(t) + \delta_X(t)$  where the time varying error  $\delta_X$  may or may not be zero-mean. The standard random uncertainty is the standard deviation of the errors  $s_{\delta_X}$ . These cannot be computed from measurements  $X$  unless  $X$  is constant, in which case  $r_X = s_X$ .

Consider the measuring of the temperature of a large object using a thermocouple. Thermocouples generate very small signals and are therefore subject to electrical noise. Meanwhile, the temperature of the large object cannot vary rapidly due to the thermal inertia.

In this case, there will likely be a large variation in the measured temperature due to electrical noise. For a single measurement, the total error is defined as the difference between the true value and the measured value. If one were to sample the temperature once, the reported value is likely to be far from the actual temperature.

This is shown in Fig. 4.1 which shows 50 samples of a thermocouple. The true temperature is 30°C, so the value of each error is the signed difference between the data point and 30°C. The value of the error of the 2nd and 5th samples are shown on the plot. Since we can be confident that the actual temperature is steady, we can estimate that  $r_T = s_T$  and compute the random uncertainty of these measurements by computing the standard deviation of the samples.

This would be the estimate of the uncertainty of each temperature sample. To determine the actual temperature of the object, it is common to average over a suitable interval of time. This can also be repeated, and each average temperature will, in general, be different from the last, although the variation will be much smaller than the difference between

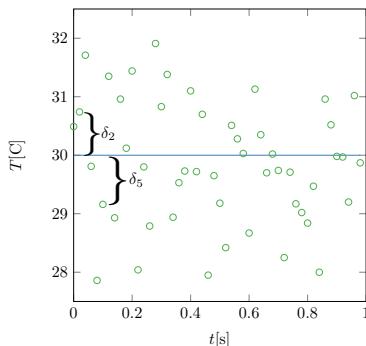


Figure 4.1: Several samples from a thermocouple that is subject to electrical noise.

the two samples. How much smaller depends on the size of the random errors and how many independent samples are acquired for each average. The uncertainty of the average temperature is a different quantity than  $s_T$  and will be discussed in Sec. 4.1.1.

Consider  $X_i$  is a discrete measurement of a fixed quantity  $X_{\text{true}}$  that is subject to the total error  $\delta_i$ . For now, we will assume that each realization of the error is independent of every other realization, which is to say that the errors are not correlated to one another, and that the error may be the net result of many error sources. Then  $X_i = X_{\text{true}} + \delta_i$ . In this case, the random uncertainty of a single sample  $X_i$  is  $s_X$ , or the estimate of the standard deviation of  $X_i$  based upon our sample. This says there is a 68% probability that  $X_{\text{true}}$  lies within  $\pm s_X$  of any  $X_i$ .

Now, consider a more complicated situation where  $X_i$  is a measurement of a time-varying quantity  $X(t)$  subject to an error  $\delta_i$ . An example is shown in Fig. 4.2. One possible goal of this experiment may be to estimate the blue curve based on the acquired green circles, which is a challenging task that could rely on the frequency of the errors being much larger than the signal. Or, perhaps the extent of variation of the signal is desired. Unfortunately, this is contaminated by noise.

Now, the variability in  $X_i$  comes from two sources: 1) Changes in  $X_{i,\text{true}}$  and 2) changes in  $\delta_i$ . It is clear that much of the variation about the mean of this signal is due to changes in  $X_{i,\text{true}}$  and that the standard deviation of the signal cannot be used to determine the random uncertainty

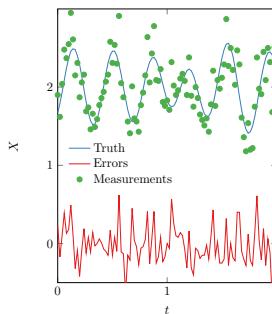


Figure 4.2: A time varying process  $X_{\text{true}}(t)$  (blue) with a mean of 2.0 that has a measurement  $\tilde{X}$  (green) that is subject to errors  $\delta$  (red) that follow a Gaussian distribution.

of the measurements. If the standard deviation of the measured signal is used as the random uncertainty, it will be overestimated. If it is used as a measure of the variability of the actual signal, it will also be an overestimate.

In some cases, the effects of the errors may be removed through filtering or phase averaging, making it possible to estimate the random uncertainty by computing the standard deviation based on the difference between the filtered and unfiltered signal. When this is not possible (e.g. when the frequencies of the blue curve and the errors are similar) the standard random uncertainty  $r_X$  must be estimated based on the instrument specifications/characteristics, such as its resolution.

#### 4.1.1 Random Uncertainty of the Mean for Independent Data

Before the uncertainty of an experiment can be assessed, the experimentalist must understand whether they are interested in the instantaneous value of a measurement or its time-average. This distinction is important because the uncertainty of these two quantities is different, with the former depending on the instrument and the latter depending on the instrument, the number of samples acquired, and the actual variation in the quantity being measured.

If we seek to estimate the fixed value  $X_{\text{true}}$  by computing the average

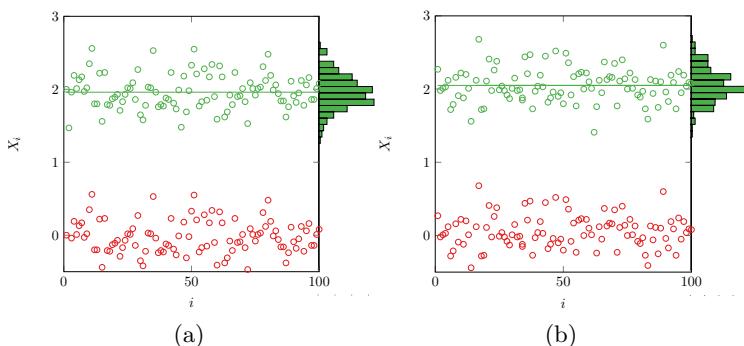


Figure 4.3: Two measurements of a fixed quantity with value 2.0 in the presence of measurement errors  $\delta$  (red) that follow a Gaussian distribution with  $\sigma = 0.25$  that is added to the measurements (resulting in the green circles). The estimate of the mean  $\bar{X}$  is shown in black. We note that the mean is underestimated in (a) and is overestimated in (b).

of  $X_i$  using Eq. 3.13, more samples (larger values of  $N$ ) should result in a better estimate of  $X_{\text{true}}$ . The quality of the estimate of  $X_{\text{true}}$  is what we call “uncertainty of the mean” and is typically expressed as a range within which we expect the true value of  $X_{\text{true}}$  to lie to some level of confidence. When  $\bar{X}$  is computed by several samples, this range is much smaller than  $s_X$ , which is the motivation for averaging.

If we repeated this measurement, a new estimate of the mean would be found (Fig. 4.3b). This shows that the estimate of the mean is also a random variable and has its own standard deviation  $s_{\bar{X}}$ . It will be shown below that the *standard deviation of the mean* depends on the variability of the measured data itself and the number of samples acquired.

For instance, we may wish to claim we have measured  $X = \bar{X} \pm s_{\bar{X}_i}$ , where  $s_{\bar{X}_i}$  is the yet-unknown standard deviation of the mean of  $X_i$ . In this case, since  $X$  is fixed, all variability in the measurements is due to varying values of the total error. Such a measurement is depicted in Fig. 4.3a where  $X_i$  is unique for each sample (due to randomly varying error) and is sampled at even intervals by a digital data acquisition system. The following discussion is equally true for samples over any independent variable and any sampling interval. If  $X_{\text{true}}$  is known to be a fixed value, the

random uncertainty of this measurement can be estimated by computing the standard deviation of the samples,  $s_X$ .

If the errors are assumed to be distributed randomly according to a Gaussian distribution, the probability of the true fixed value lying within a standard deviation of the mean of the measurements is 68%. In many fields, it is customary to express uncertainty at 95% confidence, in which case we would instead say  $X = \bar{X} \pm 1.96s_{\bar{X}_i}$ .

In this discussion thus far, we have assumed that the errors are random and uncorrelated to one another. In the parlance of other texts, the measurements are “bias free.” In our view, the requirement that the errors are uncorrelated to one another precludes the presence of an unchanging error. Correlated errors are the subject of Sec. 4.1.5.

In the authors’ experience, the fact that the mean of random data is a random variable is often difficult to digest, in spite of the fact that anyone who has performed experiments has seen it directly. It is easy to find a measurement system that will sample a sensor and report its mean after some interval. It is normal for such a system to report a different mean each time the sensor is sampled and averaged. This is a manifestation of the fact that the sample mean is a random variable.

The mean of an independently sampled random variable is itself a random variable with its own statistics. We can find this relationship by applying the Taylor Series Method to the mean operator, Eq. 3.13, which we repeat here.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (4.2)$$

For data that are independent from one another (no correlation between samples), from the TSM, the standard deviation of the mean is

$$s_{\bar{X}}^2 = \sum_{i=1}^N s_{X_i}^2 \left( \frac{\partial \bar{X}}{\partial X_i} \right)^2 \quad (4.3)$$

The derivatives are all equal to  $1/N$ . Therefore, if the uncertainties of all the samples are the same, or  $s_{X_i} = s_X$ , then

$$s_{\bar{X}} = \frac{s_X}{\sqrt{N}}. \quad (4.4)$$

The quantity  $s_{\bar{X}}$  is called the *standard random uncertainty of the mean* with “standard” meaning there is a 68% probability of the true mean  $\mu$  lying within the interval  $\pm s_{\bar{X}}$ . It is also commonly referred to as a “confidence interval” or *standard error*. If one desires a confidence interval with a 95% probability of the uncertainty bands containing the true mean, we can see from Chapter 3 that this requires 1.96 standard deviations, or  $1.96s_{\bar{X}}$ . This is commonly rounded to  $2s_{\bar{X}}$  and referred to as a “2-sigma” uncertainty.

That the uncertainty of the mean of Gaussian data is a Gaussian random variable will now be further demonstrated. The mean of a random process based on  $N$  samples will be computed  $j$  times. So each of the computed means is

$$\bar{X}_j = \frac{1}{N} \sum_{i=N(j-1)+1}^{N \cdot j} X_i. \quad (4.5)$$

In Fig. 4.4 where the same process plotted in Fig. 4.3 is sampled  $N = 25$  times and averaged, and this procedure is repeated  $j = 30$  times. Note that each of the 30 means is unique, and their distribution, as shown in the histogram, is roughly Gaussian. A histogram of the instantaneous data (the open dots) would be much wider ( $s_X > s_{\bar{X}}$ ). The uncertainty of the mean of each set of samples ( $s_X/\sqrt{25}$ ) is shown for each time-mean as an uncertainty band. We expect that these bands include the true mean (2.0 in this case) 68% of the time, and, indeed, this is true for 20 of the 30 cases. We use this to illustrate that the average of random data is also a random variable.

The standard deviation of the means shown in Fig. 4.4 (and thus the standard uncertainty of the mean) depends on the variability of the measured data and the number of samples acquired (Eq. 4.4). The standard deviation of the 30 means computed is 0.0477, while  $\sigma/\sqrt{25} = 0.05$ . The uncertainty bands are each calculated using the 30 samples to make the estimate  $s \approx \sigma$ , with  $s$  estimated from the sample of 30, so they range from 0.034 to 0.063.

In Fig. 4.5, the average of 25 samples is shown for 1000 averages, and the distribution of the means appears Gaussian. In this plot, the data (open circles) are not included for clarity. The standard deviation of the means in this case is found to be 0.048.

The mean of  $N$  samples of a Gaussian process is a Gaussian random

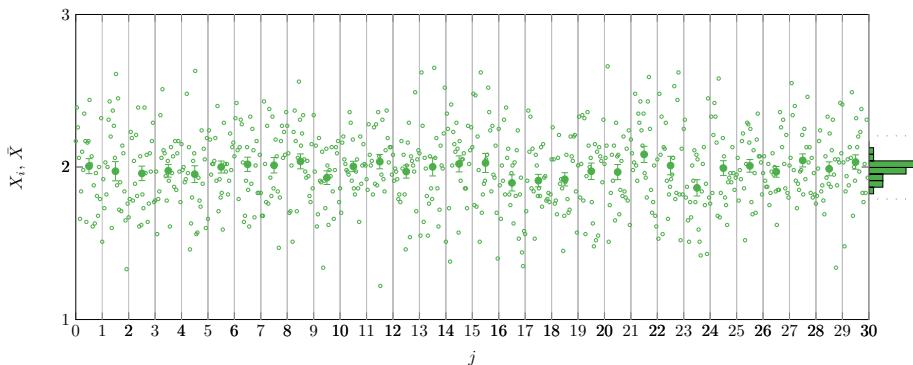


Figure 4.4: Samples (open symbols) from the same process shown in Fig. 4.3. Data are acquired for 1 second and averaged to find the time-mean (closed symbols). The standard uncertainty of the mean is plotted as an uncertainty band on the time average.

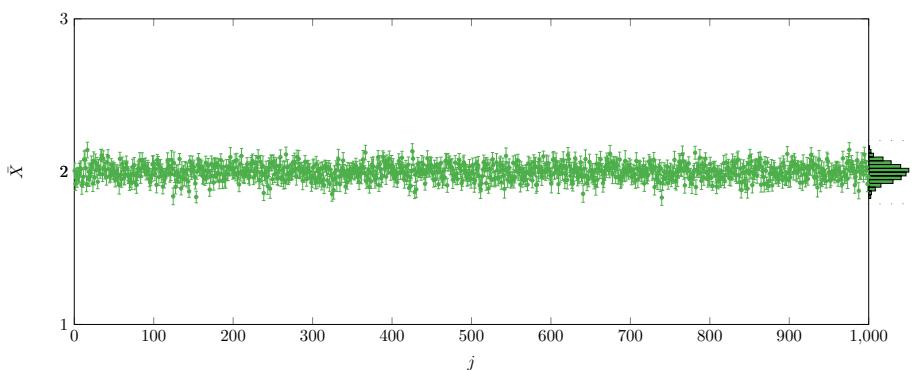


Figure 4.5: Samples means from the same process shown in Fig. 4.3. Twenty five samples are averaged to find the time-mean 1000 times. The standard uncertainty of the mean is plotted as an uncertainty band on the time average. The histogram at the right demonstrates that the mean of a Gaussian process is also a Gaussian random variable.

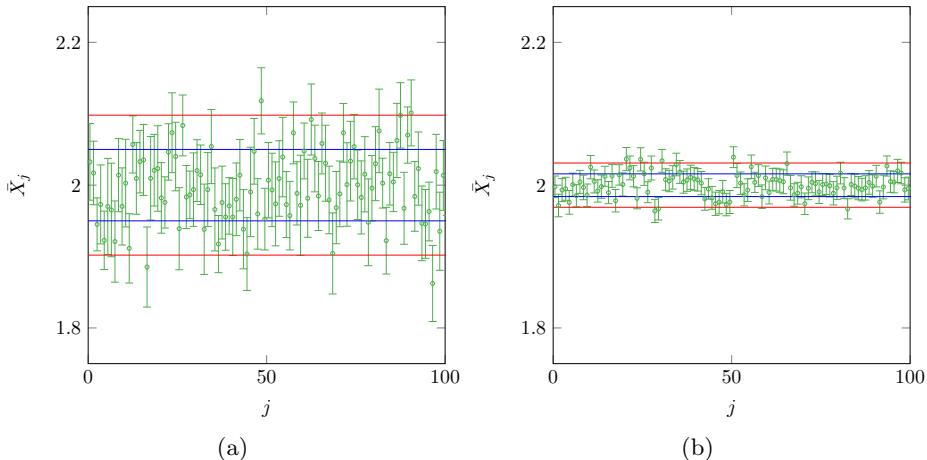


Figure 4.6: Sample means of a process with a parent mean of 2 and a standard deviation of 0.25 formed with (a)  $N = 25$  and (b)  $N = 250$ . The blue lines are  $2 \pm 0.25/\sqrt{N}$  while the red lines are  $2 \pm 1.96 \cdot 0.25/\sqrt{N}$ . The uncertainty bands are each sample standard deviation  $s_{X_j}$  divided by  $\sqrt{N}$ .

variable. The effect of  $N$  on the uncertainty of the mean as well as the Gaussian distribution of the means of Gaussian data will be further demonstrated by forming averages for  $N = 25$  and  $N = 250$ , as shown in Fig. 4.6. Note the random nature of the means. Additionally, we see approximately 68% ( $32/100$  for  $N = 25$ ) of the means outside  $s_{\bar{X}}$  and about 95% ( $4/100$  for  $N = 25$ ) outside  $1.96s_{\bar{X}}$ . If this were repeated infinite times, the number of points outside the blue lines would be exactly 68% while exactly 95% would be outside the red lines.

#### Example 4.1: Bat Speed

In Sec. 5.2 it will be shown that the standard uncertainty of a sensor used to measure the tip speed of a commercial baseball bat sensor was 0.64 mph and the standard random uncertainty, based on instrument specifications, is  $r_V = 0.22$  mph. Suppose bat speed is recorded for 50 swings ( $\bar{V} = 53.7$  mph), with the corresponding data plotted in Fig. 4.7. It is clear from the data that the swing to swing variation is much larger than the random

uncertainty (which is indicated with uncertainty bands). The question to be answered is what is the uncertainty of this mean bat speed?

First, we note that the samples are independent of one another (cf. Sec. 4.1.5). We also note that the random uncertainty of each measurement is smaller, but perhaps significant, compared to the swing-to-swing fluctuations.

The standard deviation of these samples is  $s_V = 3.6$  mph. This means, assuming the swings are a Gaussian process, that the standard deviation of the mean is  $s_{\bar{V}} = 3.6/\sqrt{50} = 0.51$  mph, which is the standard random uncertainty of the average swing speed. In other words, for a 95% confidence, the bat speed is 53.7 mph  $\pm 1.0$  mph.

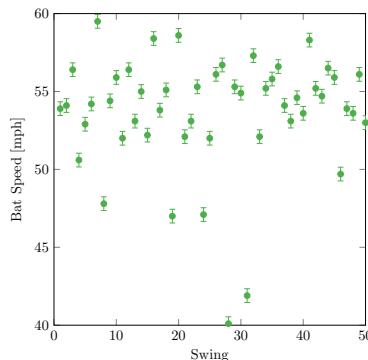


Figure 4.7: A sample of readings from a bat speed sensor. The uncertainty bands represent the expanded (95% confidence) random uncertainty of this device  $1.96r_V$  as determined in Example 5.3.

Now, consider a more complicated situation where  $X_i$  is a measurement of a time-varying quantity  $X(t)$  subject to an error  $\delta_i$ . An example is shown in Fig. 4.8. One possible goal of this experiment may be to estimate the blue curve based on the acquired green circles, which is a challenging task relying on the frequency of the errors being much larger than the signal. Or, perhaps the width of the distribution of the signal (the blue histogram) is desired. Unfortunately, this is contaminated by noise (the red histogram) to form the green histogram. Another, easier goal of the experiment may be to estimate the time-average of  $X$ ,  $\mu_X$ . For each sample,  $X_i = X_{i,\text{true}} + \delta_i$ . Now, the variability in  $X_i$  comes from two sources: 1) changes in  $X_{i,\text{true}}$  and 2) changes in  $\delta_i$ . It is clear from the

red histogram that much of the variation about the mean of this signal is due to changes in  $X_{i,\text{true}}$  and that the standard deviation of the signal cannot be used to determine the random uncertainty of the measurements. If the standard deviation of the measured signal is used as the random uncertainty, it will be overestimated. If it is used as a measure of the variability of the actual signal, it will also be an overestimate.

This issue is also common in turbulent flow measurements (where the sensor is subject to noise and the fluid velocity varies in time) and in any measurement involving living subjects. One example is the measurement of athletes performing any action repeatedly. Sensors used for these measurements are subject to error, and no athlete can perfectly repeat any task. This was shown in Fig. 4.7

In spite of the two sources of variability, the *uncertainty of the mean* is affected the same way by variations in the measured quantity as it is by errors. In other words, determining the random uncertainty of the mean requires no knowledge of the random uncertainty of the data points. One only requires the standard deviation of the data, which is impacted by both real variations and errors. In our experience, this is often quite surprising to a novice. Cases where the measured quantity varies in a repeatable way are considered in Sec. 4.1.6.

### 4.1.2 Random Uncertainty of the Mean of Small Samples

In the previous section, the concept that the mean of a random process is also a random variable with a Gaussian distribution relied on having a sufficient number of samples to calculate the mean. It is relatively easy to imagine that if the mean values shown in Fig. 4.4 were calculated from a smaller number of samples, their distribution would be wider. For fewer samples, it can be shown that the distributions of the means is wider and depends on the number of samples.

Student's  $t$  – distribution describes the distribution of mean values calculated from small samples. The variable  $t$  is defined as  $t = (\bar{X} - \mu)/s_{\bar{X}}$ , which is the number of standard deviations that the estimated mean is from the parent mean. The distribution for  $N = 6$  samples (for which the degrees of freedom  $\nu = N - 1 = 5$ ) is compared to the Gaussian distribution in Fig. 4.9a. As the number of degrees of freedom increases, Student's  $t$  – distribution becomes the Gaussian distribution. A smaller

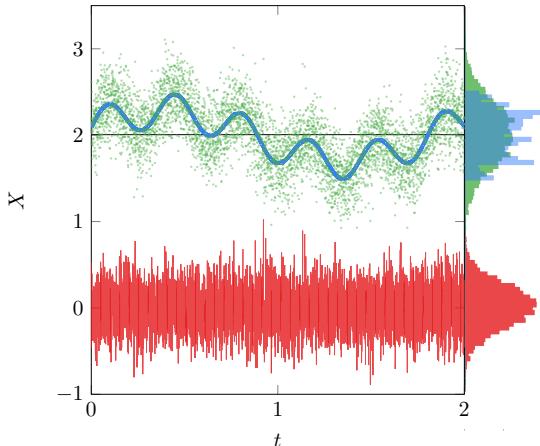


Figure 4.8: A time varying process  $X_{\text{true}}(t)$  (blue) with a mean of 2.0 that has a measurement  $\bar{X}$  (green) that is subject to errors  $\delta$  (red) that follow a Gaussian distribution. These measurements are used to compute the mean  $\bar{X}$  (black).

value of  $\nu$  will have a wider distribution and visa versa. This figure may also be used to answer the question, “How few samples is ‘small’?”. While numbers such as 30 are often cited, Fig. 4.9b shows that  $\nu = 10$  may be sufficient in many applications.

For our purpose, we only require the values of  $t$  corresponding to a 68% and 95% probability, since those are the two common confidence values used in physics and engineering. Unlike Gaussian statistics,  $t$  is a function of the number of samples as shown in Table 4.1. This table makes it clear that the mean of small samples is much closer to Gaussian at  $1 - \sigma$  (68%) than it is at  $2 - \sigma$  (95%). In other words, for 3 samples, assuming Gaussian statistics would result in an uncertainty estimate off by 30% at  $1 - \sigma$  and more than 200% at  $2 - \sigma$ .

### **Example 4.2: Finding the Rotation Rate of a Golf Ball Using High-Speed Video**

In this example, a golf ball is launched through a system that measures the velocity of the air around the ball. A specialized pneumatic cannon

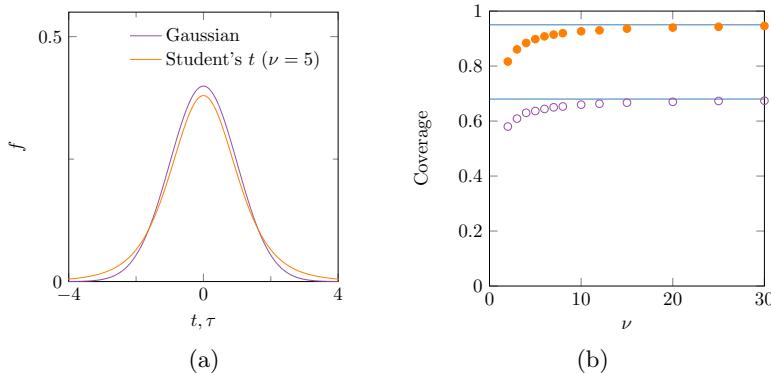


Figure 4.9: (a) The Gaussian distribution (purple) and Student's  $t$ -distribution (orange) for 6 samples and (b) The coverage obtained at 1 (purple) and 2 (orange) standard deviations as a function of degrees of freedom based on a Monte Carlo simulation, similar to as in [1]. The solid lines represent Gaussian values.

Table 4.1: Values of  $t$  corresponding to 68% and 95% coverage for various values of degrees of freedom  $\nu$ .

$\nu$	68%	95%
2	1.308	4.304
3	1.190	3.184
4	1.134	2.777
5	1.102	2.572
6	1.085	2.447
7	1.061	2.365
8	1.065	2.307
9	1.053	2.261
10	1.045	2.229



Figure 4.10: Golf ball launching cannon tip. The tip is sticky on the bottom and smooth on the top allowing it to impart spin on the ball that depends on how deep the ball is placed into the tip.

can accelerate the ball with spin due to the tip shown in Fig. 4.10. Speed is prescribed through the cannon pressure while the rotation rate depends on the depth to which the ball is placed inside the tip.

The speed and rotation rate of the ball must be set iteratively and we require a means to measure these quantities. To make this measurement, the ball passes by a high-speed camera and the number of frames required for the ball to make a full rotation is determined. If the average number of frames in one rotation of the ball is  $\bar{n}$  and the camera's frame rate is  $f$ ,

$$\omega = 2\pi f / \bar{n}. \quad (4.6)$$

For the purpose of this example, we will assume the frame rate error is insignificant compared to the ability to determine exactly 1 rotation, so no propagation of uncertainty is required (scientific high-speed cameras have very accurate frame rates). Our task is to determine the random uncertainty of  $\bar{n}$ .

The experimentalist estimates that they can find one full rotation to within 1 frame at 95% confidence. This is a measure of resolution and can be used as an *a priori* estimate (zeroth-order estimate according to [\[1\]](#)) of the uncertainty. We note that this estimate accounts only for the variability of the measurement equipment and says nothing about the variability of the actual ball rotation rate from shot to shot.

For a given setup of the launcher, the measurement is repeated 10

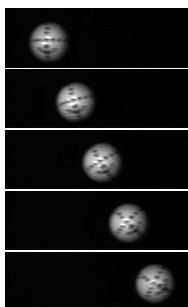


Figure 4.11: Several successive video frames of a spinning golf ball moving left to right.

times resulting in  $n_1 = 24$ ,  $n_2 = 28$ ,  $n_3 = 20$ ,  $n_4 = 24$ ,  $n_5 = 31$ ,  $n_6 = 25$ ,  $n_7 = 21$ ,  $n_8 = 30$ ,  $n_9 = 24$ ,  $n_{10} = 22$ . From these, we compute  $\bar{n} = 24.9$  and a standard deviation of 3.7. Referring to Table 4.1 for  $\nu = 9$ ,  $s_{\bar{n}} = 1.05 \times 3.7/\sqrt{10} = 1.22$ . Note that this value is larger than the experimentalist's estimate of the resolution of the measurement, and we may conclude from this that the shot to shot variation in the spin is more important to the mean value uncertainty than our ability to measure it. Acquiring more shots will reduce this uncertainty, but there is no point in attempting to reduce the random uncertainty below the resolution of the measurement (1 frame).

Assuming no correlations between the variables, the relative uncertainty of the measurement is  $1.22/24.9$  or about 5%. We note that the 1 frame *a priori* leads to a 4% relative uncertainty. However, this could be improved by viewing more rotations of the ball. If the field of view allowed 2 rotations, the *a priori* relative uncertainty would be 2%. If that number were not acceptable, more rotations could be used as long as the ball remained inside the camera field of view.

### 4.1.3 Convergence of Statistics

When random data are sampled and averaged, more data will result in a better estimate of the average. If an insufficient number of samples are used in the mean, and a second similar set of samples is used to compute the mean, two different values of the mean will result. In this case, it is common for some to say that the “results are not converged.”

The concept of random uncertainty of the mean is often convolved with convergence. It is common for experimentalists to seek convergence rather than assessing the uncertainty of the mean and requiring some small value. This is problematic for two reasons. 1) The meaning of “convergence” may not be defined and 2) for a field measurement, it is likely that any criterion for convergence will be met at one point in space for a very different number of samples than another point. This is especially true if convergence is sought in a relative sense (e.g. as a percent of the value) and the measured quantity is near zero in parts of the field. For this reason, we recommend that the degree of convergence is stated as something similar to “the random uncertainty of the mean is below 0.1 m/s everywhere in the field.”

This can be especially important when the sample standard deviation is large compared to the mean, whether that is due to noise or sample-to-sample variation of the quantity being measured. An example can be found in [12], in which the variation in the height of baseball seams from one ball to the next is described. That study showed that a sample of 72 baseballs had a mean seam height of 0.8 mm and a standard deviation of 0.08 mm. We will assume that these are parent population statistics and acquire 500 random samples from a Gaussian distribution, with these statistics, 10 times. Figure 4.12 shows the mean as a function of the number of samples for the 10 cases. The inset plot is focused on the large sample numbers and shows what many would consider a “converged” result. However, it is clear that the final mean value is different in each case. The inset plot contains a 68% (red) and 95% (blue) uncertainty band based on the standard deviation and number of samples. We hope the reader agrees, based upon this simple example, that the qualitative measure of “converged” is inferior to the qualitative statement “The random uncertainty of the mean is 0.04 mm at 95% confidence.”

#### 4.1.4 Outlier Detection

When a small sample set is used to determine the statistics of the parent population (e.g. the mean and variance), it becomes likely that one improbable sample can badly skew the result. The variance is especially sensitive to a single improbable sample.

For this reason, several schemes have been suggested over the centuries

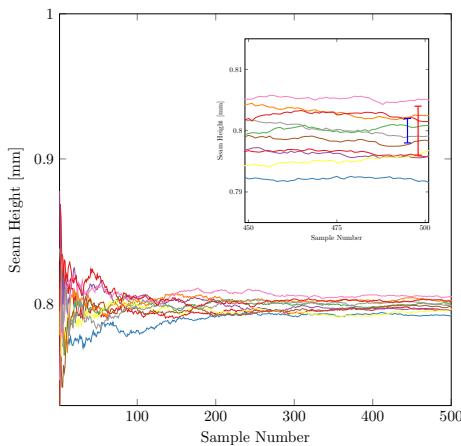


Figure 4.12: The running average of baseball seam heights with sample sizes up to 500, repeated 10 times. The inset plot is focused on the values at large sample number. A 68% (blue) and 95% (red) uncertainty band for 500 samples is included in the inset.

to detect and remove improbable samples. The most popular of these, Chauvenet's criterion was originally suggested by William Chauvenet in 1863 in an effort to simplify Benjamin Peirce's method for outlier detection.

This method has recently been proven to be ineffective [13] and detrimental to finding the population statistics from small samples. While we believe it is clear that Chauvenet's criterion should not be used to eliminate data points on a statistical basis, many [1] [2] [14] have suggested that it is a useful method for spotting illegitimate data points due to instrumentation failures. We note that it was shown in [13] that the use of Chauvenet's criterion for statistical elimination of data points is unfortunately common in the literature. This practice has considerable potential for producing the result the analyst seeks.

#### 4.1.5 Random Uncertainty of the Mean for Correlated Data

One of the biggest challenges in statistical sampling is ensuring that samples are independent of one another. In signals, we say the samples are not correlated to one another. This is important in instrumentation, but is also true of medical studies, social science, etc. A political poll cannot

be conducted in a single city or state if the result is to be applied nation wide. Furthermore, there is serious concern about the independence of polls conducted via telephone landlines, since the population that owns a landline may tend to be elderly with common views.

In order to use the analysis of the previous section, it is necessary to ensure that each sample is independent of the next. This means that if one samples in time, no waveform should be discernible. Each subsequent sample should be just as likely to be larger or smaller than the last. An example is shown in Fig. 4.13 where a hot wire anemometer is sampled for 0.1 seconds. Sampling at a high-rate allows one to observe the time-behavior of velocity, and the passage of turbulent structures in the flow is evident in the chaotic signal with a period near 0.001s.

Achieving independent sampling requires that the sample period be much longer than the largest period present in the signal. We note that this is the opposite of satisfying the Nyquist criterion (sampling at a minimum of twice the highest frequency, [15]). Sampling independently not only precludes resolving the largest frequency in the signal; it also precludes resolving *any* frequency. If the aim of the sampling is to determine frequencies in the signal, independent sampling becomes impossible. As we will show, the extent to which data are independent is captured in the autocorrelation of the signal.

It is also clear in Fig 4.13 that acquiring the same number of independent samples requires much more time than sampling at larger frequencies. This may present significant costs. Additionally, as data acquisition hardware (whether analog-to-digital converters or cameras) become faster, more diligence is required to avoid samples that are correlated to one another. Putting this another way, if the same number of samples of the waveform in Fig 4.13 are acquired at 250 or 15,000 samples per second, the error between the sampled mean and the population mean is likely much larger for the samples at 15,000 Hz, even though  $s_u/\sqrt{N}$  is the same.

For samples that are positively correlated, convergence of the mean happens more slowly than Eq. 4.4 predicts. Since one needs to control the random uncertainty of the mean by choosing the sample number, it is important to be able to estimate this convergence rate, even for dependent samples.

It has been shown by Sciacchitano and Wieneke [16] that the variance

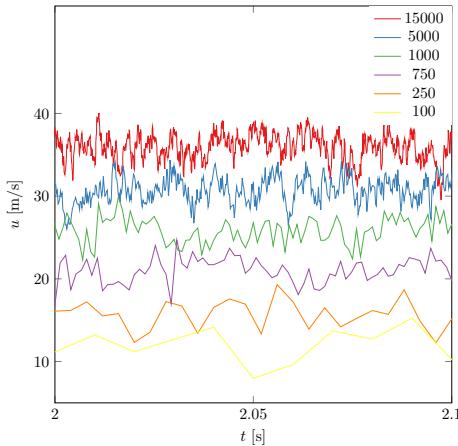


Figure 4.13: Samples from a hot-wire anemometer system acquired over the same length of time but at different sampling rates, shown in the legend.

of the mean of correlated data is larger than that predicted by Eq. 4.4,

$$s_{\bar{X}}^2 \approx \frac{s_X^2}{N} \left( 1 + 2 \sum_{j=1}^k \hat{\rho}_j \right). \quad (4.7)$$

where  $\hat{\rho}_j$  is the biased sample autocorrelation function,

$$\hat{\rho}_j = \frac{\frac{1}{N} \sum_{i=1}^{N-|j|} (X_i - \bar{X})(X_{i+|j|} - \bar{X})}{s_X^2}. \quad (4.8)$$

The autocorrelation is identically 1 at  $j = 0$  and generally goes to zero for large values of  $j$ . The more correlated samples are, the slower the decay. For independent samples, the autocorrelation will be near zero for any positive value of  $j$  and Eq. 4.7 reduces to Eq. 4.4.

Sciacchitano and Wieneke took the unique approach of propagating random uncertainty through the definition of the time-mean using a Taylor Series Method [1] including the correlated terms that involve the autocorrelation of the samples. Another way of viewing the same results is an

“effective” number of samples should be used in Eq. 4.4, and that the number of effective samples is generally less than the number of actual samples when the data are not independently sampled. The scaling depends on the autocorrelation,

$$N_{eff} = \frac{N}{1 + 2 \sum_{i=1}^k \hat{\rho}_j} = \frac{N}{\lambda}, \quad (4.9)$$

where  $N$  is the number of samples acquired. The quantity  $\lambda = 1 + 2 \sum_{i=1}^k \hat{\rho}_j$  happens to be half the integral time scale used in fluids literature [17]. These authors suggest, as is common, to terminate the summation when the autocorrelation first crosses zero rather than summing over all  $k$  lags.

Smith *et al.* [11] discussed issues with this approach. Accurate estimation of  $\rho$  requires ensemble averaging of many data sets, each of which must be many integral scales long. A sufficiently large set for this calculation will likely have a very small uncertainty of the mean, making the process largely meaningless.

Smith *et al.* also showed that the integration of  $\rho$  over all lags based on a single record is identically zero. They compared the efficacy of truncating the integral of  $\rho$  at the first zero crossing or at its minimum and showed that the best method depends on the shape of the autocorrelation, which cannot be known *a priori*.

As a result, approaches based on “bootstrapping” methods are recommended [11]. Bootstrapping estimates the standard deviation of the mean of a limited set of data by subsampling that data randomly and computing the mean from that subsample. The standard deviation of many such subsample means is used as an approximation of the standard deviation of the mean. There are many different bootstrapping techniques that use different methods to choose the “block” of data to be average as well as the length of that block. Interestingly, the effect of some of these methods can be reduced to a closed-form expression involving mostly the same quantities as Eq. 4.7.

The application of the stationary bootstrap to the computation of the uncertainty of the sample mean does not require computer-based resampling. The stationary bootstrap [18] produces a formula for the uncertainty

of the sample mean involving a weighted sum of the autocorrelation coefficients, with weights dependent on the block length  $b$ .

$$s_{\bar{X}}^2 \approx \frac{s_X^2}{N} \left( 1 + 2 \sum_{j=1}^{N-1} b_j \hat{\rho}_j \right), \quad (4.10)$$

where

$$b_j = \left( 1 - \frac{j}{N} \right) \left( 1 - \frac{1}{b} \right)^j + \frac{j}{N} \left( 1 - \frac{1}{b} \right)^{N-j}. \quad (4.11)$$

In other words, resampling using this method and estimating the standard deviation of the mean based on the resampling will give the same result, on average, as Eq. 4.10. One must still choose the block length  $b$ . For bootstrapping methods, the work of Politis and White [19] provided an *automatic* selection algorithm for choosing the block length  $b$  for either the moving block or stationary bootstrap methods based on the flat-top lag window for spectral estimation. That is, the bootstrap methods have an additional virtue of accompanying algorithms in the statistics literature which remove the human element from parameter selection in their application. While other bootstrap methods were considered, these authors concluded that, absent any *a priori* information about the autocorrelation, for a single record, stationary bootstrap gives the most reasonable and stable estimate.

#### 4.1.6 Phase and Conditional Averaging

In some cases, a measurement output may repeat periodically. If a signal contains a known, dominant frequency, it may be advantageous to average each point of a cycle over many cycles, or what is known as “Phase Averaging.” This works best when the experiment is driven with a known period so that samples may be acquired at a subinterval of that period. If the period is not enforced but occurs naturally, such as with periodic vortex shedding from a bluff body, conditional averaging, discussed below, is more appropriate.

Data from a phase-locked experiment is shown in Fig. 4.14. The sampling period is 1/100 the forcing period, so 100 samples are acquired per cycle. There is significant variation from one cycle to the next, which may be due to cycle-to-cycle variations in the flow or non-repeatability of

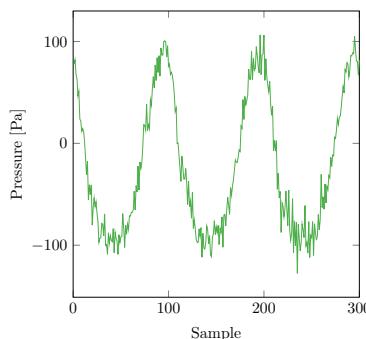


Figure 4.14: Several cycles of phase-locked pressure measurements made on the wall below a diffuser with oscillating flow [20].

the pressure sensor. The sensor specification puts the non-repeatability at 13 Pa, which appears to be in line with the variations seen.

By averaging all 100 cycles together, the effect of sensor non-repeatability and cycle-to-cycle fluctuations can be mitigated. Fig. 4.15 shows the result of phase-averaging the pressure data from Fig. 4.14. The sensor specified “non-repeatability” is a good estimate of the random uncertainty of the sensor as described in Sec. 4.2.4. A standard deviation can be computed at each phase point using each cycle’s pressure value and the phase average. This is shown in red in Fig. 4.15. We note that the value is nearly the same at every phase and is very close to 13 Pa. We therefore conclude that the bulk of the “noise” in Fig. 4.15 comes from the sensor and not cycle-to-cycle variation in the flow.

The uncertainty of the phase averaged pressure can be determined in the same manner as any average. Assuming that each cycle is independent of all others, the random uncertainty of the phase average is the standard deviation divided by the square root of the number of cycles, which is a value near 1 Pa. This uncertainty is manifest in Fig. 4.15 in the deviations from a smooth curve that are about 1 Pa in size.

As discussed above, if the system under study is not forced at a known frequency, it may still be desirable to perform something similar to phase averaging. However, without access to a driving waveform, it is not possible to precisely know the period of the waveform. An alternative to phase averaging is called “conditional averaging”. In this scheme, some distinctive

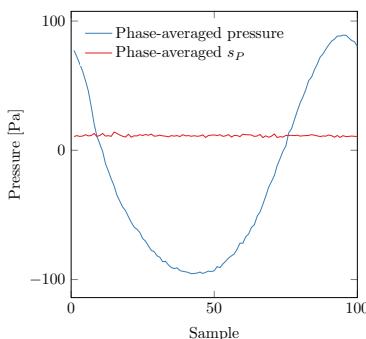


Figure 4.15: Phase-average of 100 cycles similar to those shown in Fig. 4.14.

event is used as a “trigger” and becomes the starting point of the sample. Successive points at even intervals are averaged over many ensembles. As one samples farther and farther from the trigger event, cycle to cycle variation will come to dominate and the averaged signal will converge to the time average.

## 4.2 Uncertainty of a Single Variable From Known Sources

Systematic uncertainties are the result of errors that do not change when additional samples are acquired. In many cases, the error source varies randomly among identical pieces of equipment (e.g., measurement systems, sensors). However, a single piece of that equipment, meaning a single sample of the calibration error, may be used in an experiment.

As described below, many uncertainties are a result of the calibration procedure (e.g. the calibration source uncertainty). Uncertainty can come from how the calibration result is modeled (e.g. non-linearity uncertainty).

A lack of spatial resolution can lead to errors. Describing these as uncertainties can be very challenging, since the lack of spatial resolution makes it impossible to know what spatial resolution is required.

### 4.2.1 Uncertainty From Reading Analog Devices

Many measurement devices employ an analog readout of some kind. Some of the most obvious examples are rulers and thermometers. Like any measurement device, these instruments suffer from errors due to calibration which will be constant over short times. In addition to these, there are additional errors that stem from the user reading the value. While one user may be biased toward the high or low side of the scale, we wish to focus on any user's ability to interpolate between two divisions on the scale. It is customary to assume that a well-constructed scale will result in errors leading to a total (95% coverage) random uncertainty equal to 1/2 of the smallest division [1]. When choosing the spacing between divisions, a manufacturer must consider 1) the user's ability to discriminate between the divisions and 2) the calibration uncertainty of the device. We recommend assuming that the total standard uncertainty of these devices is 1/2 of the smallest division and that this stems entirely from random uncertainty.

An example of such a device is shown in Fig. 4.16. This is a mercury manometer used as a pressure calibration source. By assuming a standard uncertainty of this device of 0.05 inches Hg, we are assuming that the root sum squared of its calibration error plus the error associated with reading the device will result in a distribution of errors with a standard deviation of 0.05 inches Hg.

Some instruments have a digital display that must be read by a user and recorded. For such devices, a random uncertainty equal to half of the last digit is a reasonable assumption, although we caution that displays are much cheaper than sensors, and some unscrupulous manufacturers may employ more digits than are sensible.

### 4.2.2 Uncertainty Due to an Error Source With a Periodic Input

It is common for instruments to have sensitivity to uncontrolled variables, with ambient temperature being the most common. One will often find the temperature sensitivity of a device on an instrument specification sheet. Several examples are shown below in Sec. 4.2.4.

In the case where a sensor has an error proportional to ambient temperature, the impact of this error depends on how temperature varies. We



Figure 4.16: A mercury manometer as an example of an analog device. Scale divisions are 0.1 inch Hg.

will focus on two common patterns: a sinusoidal variation (which can be a good model for outdoor temperature) and a sawtooth wave (which is a good model for indoor spaces with conventional HVAC equipment).

In either case, one must estimate the extent of temperature variation. For the indoors case, this is usually set by the room thermostat. It is a simple matter to find data on typical outdoor locations given the location and time of year.

Consider a case where a laboratory thermostat controls the heating of a room between the limits of  $T_L = 20^\circ\text{C}$  and  $T_H = 24^\circ\text{C}$  as shown in

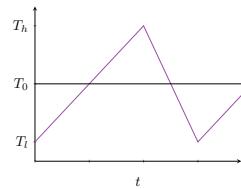


Figure 4.17: Room temperature variation over time as controlled by a typical HVAC system.

Fig. 4.17. We assume 1) a sawtooth wave behavior of temperature, 2) the resultant errors are proportional to the ambient temperature, so they also have a sawtooth wave behavior and 3) that any error due on the average temperature of the room  $T_0$  has been removed. If the temperature sensitivity of the device is denoted as  $\beta$  (with units of Error/degree), then the error sawtooth will range between  $\beta T_h$  and  $\beta T_l$ . In Sec. 3.4.2, the PDF of a sawtooth wave is uniform with  $f_y = 1/(y_u - y_l)$ , or in this case,  $f_{\beta T} = 1/\beta(T_u - T_l)$ . We know from Sec. 3.3 that the standard deviation of the uniform distribution is  $\beta\Delta T/(2\sqrt{3})$  where  $\Delta T$  is the difference between the upper and lower temperature. So, the uncertainty from this error source is  $u_{X_T} = \beta\Delta T/(2\sqrt{3})$ .

Now let's consider a sinusoidal temperature variation.

### 4.2.3 Instrument Calibration

Nearly any instrument requires some type of calibration. Most users rely on the instrument manufacturer or a calibration facility to perform the calibration. This section is included for two reasons: 1) to improve the reader's understanding of the calibration process so they can better understand the uncertainty specifications that come from it and 2) to encourage readers to consider performing their own calibrations.

#### Understanding the Calibration Procedure

Calibration involves relating the output of an instrument, usually a transducer (meaning a sensor coupled to circuitry to generate a voltage output) to a change in the quantity to be measured. This requires

1. A “calibration source.”
2. A means to alter the quantity of interest
3. A transducer sensitive to that quantity

A *calibration source* is the instrument that is compared to the output of a transducer to form the relationship between the transducer output and the quantity being measured. Ideally, a calibration source has a performance that does not vary over time and therefore cannot be “out of calibration.” Often, portability is sacrificed in order to achieve the more important feature of performance stability.

An excellent example is a mercury manometer. This is a device which must be mounted vertically on a wall. Using mercury as the fluid provides stability, as its density is not prone to change over time. The device relies only on our knowledge of the density of mercury and the local acceleration due to gravity, each of which is well known. For many, these benefits offset the potential hazards and regulatory requirements that come with large amounts of mercury in a laboratory.

Platinum Resistance Temperature Devices (RTDs) offer the similar benefit of stability over time but can be fragile, expensive, and require specialized and expensive circuitry to read. As a result, they are often used as a temperature calibration source.

Calibration requires a means to vary the quantity of interest. For pressure calibration of relatively low pressures in air, a small pressure vessel with an air pump can suit this purpose. For temperature calibration, a small furnace can be used. In some cases, the first two requirements are combined in a single unit, such as a thermocouple calibrator, which is commercially available.

To calibrate, the quantity of interest is set to a value within the desired calibration range and the calibration source and transducer are read simultaneously. Several pairs of calibration source values and transducer outputs are recorded for different values of the quantity of interest. The number of points required depends on the random uncertainty of the calibration source as well as the transducer response. Linear transducers require fewer points than non linear transducers.

An example of a pressure calibration is shown in Fig. 4.18 a. In this case, there is significant error associated with reading the manometer (cf. Sec. 4.2.1), and many data points are acquired to mitigate the effect of these errors. The result of the calibration is the equation of the blue line,  $P = -0.182 + 4.39V$ , where  $V$  is the transducer output voltage and pressure is in units of inch Hg.

The nature of the errors from reading the analog scale may be examined by plotting their histogram, shown in Fig. 4.18 b. There is a strong central tendency and it appears plausible that these errors are drawn from a Gaussian pdf. Additionally, we note that the width of this distribution is roughly 1/2 the smallest division on the manometer. This fits with guidance provided by [1] as discussed in Sec. 4.2.1.

This pressure transducer has the desirable feature of a high degree

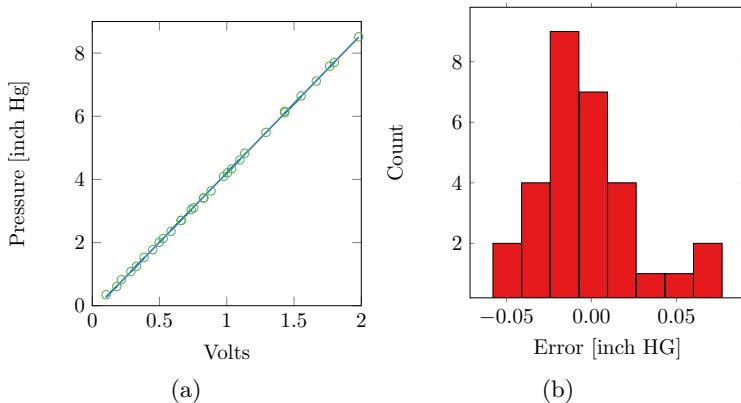


Figure 4.18: (a) Calibration data of a pressure transducer against the manometer shown in Fig. 4.16. The independent variable is the voltage output of the transducer. (b) Histogram of the difference between the data points and the line fit showing that errors from reading an analog device can be modeled as Gaussian distributed with a width of approximately 1/2 of the smallest division on the scale.

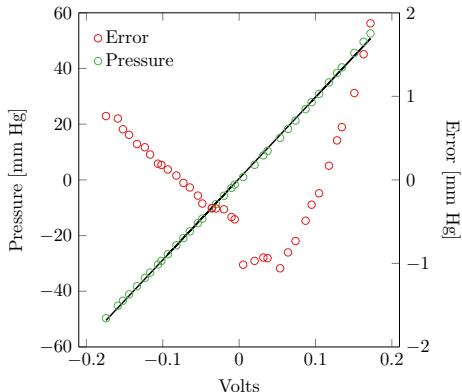


Figure 4.19: Calibration data for a piezoresistive pressure sensor that exhibits nonlinear behavior. The error between the line fit and the calibration data is shown in red symbols.

of linearity over the calibration range, and therefore fits the line well. Many transducers have a repeatable range that exceeds their linear range, and using a linear fit for these transducers results in an error termed “non-linearity error.” An example of calibration data for a different, piezoresistive pressure sensor is shown in Fig. 4.19. Note how the linear fit is significantly below the data for large negative and positive pressures and above the calibration data near zero. The calibration range in this case is beyond that for which the sensor remains linear. Sensors like this are frequently sold with a linear calibration and a specification for non-linearity error. This particular device has a specification of nonlinearity error that is “typically” less than 1% of full scale output of 52 mm Hg. Clearly, the error becomes much larger than this near the full scale output.

We note that this error is only incurred because we chose to assume that the transducer was linear when it was not. The calibration data will easily fit a polynomial, but most manufacturers do not provide that fit to the data, and in-house calibration would be required to find the polynomial constants.

## Benefits of In-House Calibration

Many laboratories lack the ability to calibrate their own sensors despite the fact that, in at least some cases, calibration equipment is not prohibitively expensive. We believe this may be due to a lack of understanding of the calibration process or a lack of appreciation for the many benefits of this capability.

In-house calibration allows one to be aware of departure from calibration. This is especially important if something has occurred that makes one concerned that the calibration may no longer be valid. For instance, consider a case where a scale has been over-ranged by 20%. One would hope that the calibration would not be altered, but to be sure requires shipping the device to a calibration facility and waiting for its return.

In addition to this obvious benefit, there are also several less obvious benefits of in-house calibration. The first of these is that the calibration range or procedure may be more closely aligned to the experimentalists needs. In-situ calibration eliminates installation errors. Additionally, some manufacturers limit their calibration range in order to limit non-linearity errors. If the user is aware of the non-linearity and can use a more flexible calibration function (e.g a polynomial), then the calibration range can be extended while eliminating this error source. Using a standard thermocouple response function assumes that the user's sensor has an identical composition to the NIST standard. Calibrating them in-house eliminates the error due to assuming an ideal thermocouple, which can be more than 2°F.

Perhaps the least obvious benefit of in-house calibration is to gain control over correlated errors between sensors. As will be shown in Chapter 5, measurements based on multiple sensors can have uncertainties that are either larger or smaller depending on the degree of correlation between the multiple measurements. If multiple sensors are calibrated against the same calibration source, that source becomes a correlated error between the sensors.

### 4.2.4 Using Instrument Specifications For Uncertainties

Instrument specification sheets can be confusing. Unfortunately, the way that manufacturers write instrument specifications is non-standard, uses inconsistent terminology, and can be very difficult to relate to uncertainties.

Some of this stems from the fact that not all sensors suffer from the same issues. For instance, sensors that have linear response will not report nonlinearity specifications. Sensors that are sold without electronics will most likely say nothing about their noise levels. This section will provide some guidance on how to use instrument specifications.

One shortcoming that nearly every instrument specification sheet the authors have found has in common is a lack of information about the assumed distribution that errors follow and what confidence interval is reported. Absent this information, one must assume that manufacturers assume a Gaussian error distribution and that  $2 - \sigma$  (95% confidence) values are reported. It is unfortunately common for manufacturers to report a maximum error rather than an uncertainty. While this may be useful to a user who is using only that instrument, it is not possible to propagate that error result into a final result based on measurements of several quantities.

## Absolute Versus Relative Uncertainty

Some sensors have the property that their errors, and thus uncertainties, scale with the reading. For such sensors, it is common to specify their uncertainty as a percentage of the reading. It is important to note that this is a dimensionless quantity. The dimensional uncertainty of a sensor that measures  $X$  and has an uncertainty of 2% of reading is  $u_X = 0.02X$ .

For sensors with a fixed uncertainty, it is common to list the specifications as a percentage of the full scale output (FSO). Note that FSO is fixed, so these values do not change with reading and the uncertainty has units. These devices are generally inferior to those whose uncertainty is a function of reading, and great care must be taken when using them at the low end of their scale.

Unfortunately, in the authors' experience, specification sheets very commonly have entries that are labeled as a percentage without stating whether this is of reading or full scale output. In this circumstance, one must rely on one's knowledge of the measurement technology (e.g. piezo-resistive elements have a fixed accuracy) or contact the manufacturer.

## Hysteresis

Hysteresis is the tendency of a sensor to have a response that is dependent on the direction in which the input variable is changing. In other words, the sensor will report different values for the same input depending on if the input is increasing or decreasing. This is common among strain-gage-based sensors including pressure and force sensors. How this tendency impacts a measurement depends on how the sensor input varies. If one has no information on how the input will vary, this should be assumed to be an additional source of random uncertainty.

## Non-Repeatability

Non-repeatability specifies the standard deviation of several reported values under the same condition but arrived at randomly and independently [L4]. If an instrument is used to measure a randomly varying quantity, or even for an oscillating quantity, non-repeatability can form an important part of the random uncertainty.

## Sensitivity

Sensitivity is the relationship between a change in input and a change in output. In most cases, sensitivity is the calibration constant of the instrument. It is often provided on the specifications of a sales brochure of an instrument in a broad range, while a specific sensitivity based on the calibration of a specific sensor will be included with the unit when shipped. This information is useful for propagating limits of voltage resolution of the data acquisition system.

## Accuracy

Accuracy is a term commonly used in specification sheets to indicate the uncertainty of a device. Most authors agree that it is an expression of the difference between the measured and true value (an expression of error) and differentiate it from random uncertainty. Some caution should be applied that a specification may indicate maximum error rather than uncertainty. An accuracy expressed as a negative number is a clear case where it is an expression of error rather than uncertainty.

## Precision

Precision is a term that is often used in uncertainty literature to convey the repeatability of measurements. A description of precision is often accompanied with a series of bulls eye targets and discussions of "accuracy vs. precision" that are visually represented by showing arrows fired at the bulls eye. The arrows are meant to represent the measurements and the bulls eye is said to represent the true value. This description will usually imply that if there are a tight cluster of arrows, then that these measurements are more precise, regardless of how they compare with the true value. However, this is very similar to what has been established as a random error and using a term such as precision only serves to cause more confusion. Using a term such as precision interchangeably with bias error is additionally problematic since precision may also refer to the number of digits to which a value can be measured, i.e. a value known to larger number of digits is also said to be more precise. The authors therefore advise against using the analogy of the bulls eye with the terms "accuracy and precision" and instead use the terms "random" and "bias" in regards to their associated errors and uncertainties.

## Resolution

Resolution is the minimum detectable change in the measured quantity. If no larger error source exists, this can be used as a random uncertainty value.

## Nonlinearity Error

We saw in Sec. 4.2.3 that assuming a linear transducer response in spite of non-linear behavior produces an error. The nonlinearity errors (the difference between the line fit and each data point) shown in Fig. 4.19 are plotted in histogram form in Fig. 4.20. The standard deviation of these errors is 0.7 mm Hg. Recall that the manufacturer's specification on "typical" nonlinearity error for this device was 0.5 mm Hg. So while a "typical" error may be difficult to convert to 68% coverage, at least in this case, they are nearly the same, assuming the full range of the sensor was being used in a somewhat random manner.

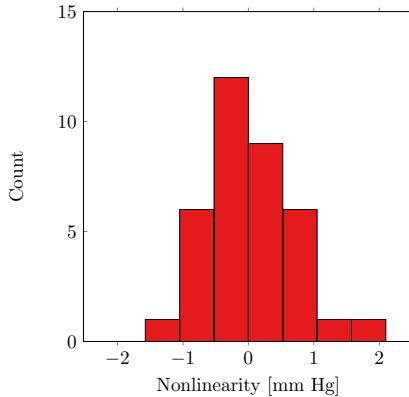


Figure 4.20: Histogram of the nonlinearity errors (the difference between the line fit and each data point) from Fig. 4.19.

### Irrelevant Specifications

It is common for instrument manufacturers to provide specifications that indicate a very high accuracy without acknowledging the significant installation errors that cannot be avoided. For example, the error of a heat flux sensor is dominated by installation errors (mismatch of thermal resistance which may cause the gage to alter heat flow, pressure on the gage that may alter its thickness, etc.).

In other cases, the data reduction equation (DRE) being used may contain variables that are difficult to quantify. Infrared cameras used for temperature measurements operate on a DRE that is based on the Stefan–Boltzmann law.

$$q'' = \epsilon\sigma T_s^4 \quad (4.12)$$

where  $q''$  is the radiative flux,  $\sigma$  is the Stefan–Boltzmann constant,  $\epsilon$  is the surface property emissivity, and  $T_s$  is the temperature of the surface being measured. Cameras specifications normally relate to  $q''$ , which can be measured accurately. However, in order to interpret the flux as temperature,  $\epsilon$  must also be known. The DRE is  $T_s = (q''/\epsilon\sigma)^{1/4}$ .

Unfortunately,  $\epsilon$  (which is defined as a value between 0 and 1) is very

difficult to measure and varies with temperature, angle of incidence, and condition of the surface. In order to use a thermal camera to measure temperature, one typically enters a constant value of  $\epsilon$  for the entire field. This value quickly becomes the largest source of uncertainty in this measurement, making the instrument specifications irrelevant. In such cases, it is unfortunately common practice to quote the instrument specifications rather than doing the hard work of determining the impact of the unknown variables.

#### 4.2.5 Examples of Use of Instrument Specification

This section will provide examples of how to form an uncertainty from the specification sheets of a few common instruments. While we will not identify the instrument manufacturers by name for obvious reasons, these are actual commercially available sensors. These examples are by no means meant to be exhaustive. Hopefully, taken together, they can aid the reader in determining uncertainties for any sensor.

#### Piezoresistive Pressure Sensors

The first sensor is a 1-PSI Full Scale Output (FSO) device. No accuracy is reported, but instead several different error sources are specified as a percentage of FSO. **Nonlinearity** is 1%, while **non-repeatability** and **hysteresis** are 0.2% of FSO. This sensor is sold without electronics, so the noise added to the reading is not specified.

A second device, from a different manufacturer and which is sold with electronics, uses an electrical noise specification as a resolution. The full specifications for a 50 psi unit are: **resolution** 0.0005 psi, (rms) **nonlinearity** of  $\pm 1\%$  Best Fit Straight Line (BFSL), and **hysteresis** of 1%.

Hysteresis or non-repeatability errors generate uncertainties depending on the measurement being made. For instance, if a steady pressure is being measured, hysteresis and non-repeatability will generate an error that does not change in time and depends on the value of pressure being measured in a manner that is not well known. However, if the pressure is variable, the input to this error (the pressure value itself) is variable and the effect will be quantified through random uncertainty. We note that this error distribution is unlikely to be Gaussian and will be bounded.

We also note that the working principle of these devices, a full-bridge formed with silicon strain gages, is used in other sensor types, such as load cells. These devices will have similar specifications.

### Capacitive Pressure Sensors

This device is specified with a **range** of 1 mm Hg, a **resolution** of  $1 \times 10^{-6}$  mm Hg, an **accuracy** of 0.12% of the pressure, and a **usable range** from  $2 \times 10^{-5}$  to 1 mm Hg. Of these specifications, the most important is accuracy, which we will take to be twice the standard uncertainty  $2u_P = 0.0012P$ . The sensor resolution should be used as twice the random uncertainty,  $2r_P = 1 \times 10^{-6}$  mm Hg. This will only become significant if the device is being used over a very narrow range of pressures and is amplified to the point that the DAQ resolution does not become the larger issue.

### Turbine Flow Meter

A certain model's standard unit has an **accuracy** ("including **linearity**, best fit straight line") of  $\pm 1.0\%$  of full scale output and a **repeatability** of  $\pm 0.2\%$ . Additionally, the specifications list a temperature sensitivity  $\pm 0.2\%$  FSO or less per Celsius degree. More expensive versions of the same unit have better specifications.

In this case, we would use the accuracy specification as an uncertainty. If the repeatability specification is concerning, sampling should be used to quantify its impact through a random uncertainty. However, with the repeatability being 1/5th the accuracy, this is clearly an insignificant uncertainty.

But, what if the fluid being measured has a variable temperature? If a  $5^\circ\text{C}$  random variation of fluid temperature is assumed, this would form a random uncertainty of a similar size to the uncertainty due to accuracy.

### Thermocouples

Thermocouples are ubiquitous, most likely because they are cheap and robust. However, they have rather large uncertainties compared to other temperature measuring devices that stem from several sources. The specifications for these may appear in different documentation since some are

related to the sensors themselves while others may be related to the analog to digital electronics. We will discuss these separately.

Thermocouples are made from wires of two different alloys, and the uncertainty of the composition of these alloys leads to uncertainty in their response. However, the standard response functions for each type of thermocouple assume that the alloys are as specified. Deviation from this standard is quantified by an uncertainty specific for each thermocouple type. This is usually specified two ways, and the larger of these two values is taken as  $2u_T$ . For instance, one manufacturer of T-type thermocouples specifies that they are accurate to  $1.0^\circ\text{C}$  or  $0.75\%$  above  $0^\circ\text{C}$ .

This uncertainty can be completely removed through calibration, in which case, it should be replaced with the uncertainty of the calibration source. If the DRE involves differences in temperatures (which is common), calibrating the thermocouples in the same facility will cause the calibration source error to be correlated, which reduces uncertainty, as discussed in Chapter 5.

Additionally, the analog-to-digital conversion equipment introduces additional uncertainty. Typically, the largest source of error in this equipment is the “cold-junction temperature” (CJC) sensor. If thermocouples had a linear response, any error on the CJC would be added to the error on the thermocouple temperature. In reality, thermocouples are not completely linear. Additionally, the thermocouple junctions may not be at the CJC temperature due to heating from other nearby electronics. Nevertheless, we recommend treating the CJC uncertainty as an uncertainty on the measured temperature.

As an example of computing the total uncertainty of a thermocouple measurement, consider a measurement system that has a T-type thermocouple and measurements are being made near  $50^\circ\text{C}$ . A total of  $N = 100$  samples are acquired resulting in  $s_T = 1^\circ\text{C}$ . We seek the uncertainty of the average temperature, so  $r_{\bar{T}} = s_T/\sqrt{N} = 0.1^\circ\text{C}$ . An uncertainty from the CJC is specified as  $u_{T_{CJC}} = 1.5^\circ\text{C}$  and the uncertainty of the sensor itself is  $u_{T_{TC}} = 1.0^\circ\text{C}$ . Therefore the total standard uncertainty of the average temperature measurement is

$$u_{\bar{T}} = \sqrt{r_{\bar{T}}^2 + u_{T_{CJC}}^2 + u_{T_{TC}}^2} = 1.8^\circ. \quad (4.13)$$

### 4.2.6 Uncertainty From Material Property Data

In their influential textbook, Coleman and Steele [1] correctly point out the importance of material property data in uncertainty analysis. In the author's experience, students are prone to assuming property data are free from error, perhaps because the tabulated data appear smooth and their origin is obscure. It is common for property data in thermodynamics textbooks to be generated by software using curve fits to the original data, completely obscuring any issues with the original data.

Many property values were originally measured many decades ago, and their source can be difficult to find. The most accessible source is NIST (formerly the National Standards Bureau). An example, the thermal conductivity of aluminum, is shown in Fig. 4.21. The results of dozens of studies are shown, and it is clear that the exact values remain somewhat controversial. When faced with determining the uncertainty of a quantity such as thermal conductivity of aluminum, one is faced with several poor choices. One could treat each study as an independent sample and attempt to estimate the standard deviation of all values reported at one temperature, as advocated by [1], or one could use the guidance provided in [21], which points to a specific results and provides some rather vague recommendation ("The recommended values are thought to be accurate to within  $\pm 4\%$  below room temperature and  $\pm 2\%$  to  $\pm 3\%$  above. For liquid aluminum the values are probably good to within  $\pm 5\%$ ").

We note that while many of our instruments improve over time, these data do not. For instance, the thermal conductivity of water is still not known to be much better than 2% and can be as bad as 5% [22]. It is unlikely that anyone will ever produce new measurements on the thermal conductivity of water again. Furthermore, in many cases, the uncertainty on a material property may be due more to errors in the material composition (e.g. alloys) than in the measurements. As time goes on, it becomes increasingly likely that uncertainty from material properties can dominate our overall uncertainty.

### 4.2.7 Uncertainty of Coefficients of a Polynomial Fit

It is common in measurements to use the result of a curve fit. Perhaps the most obvious example is a calibration where the result is the slope of the input/output relationship. We will begin with the more general case

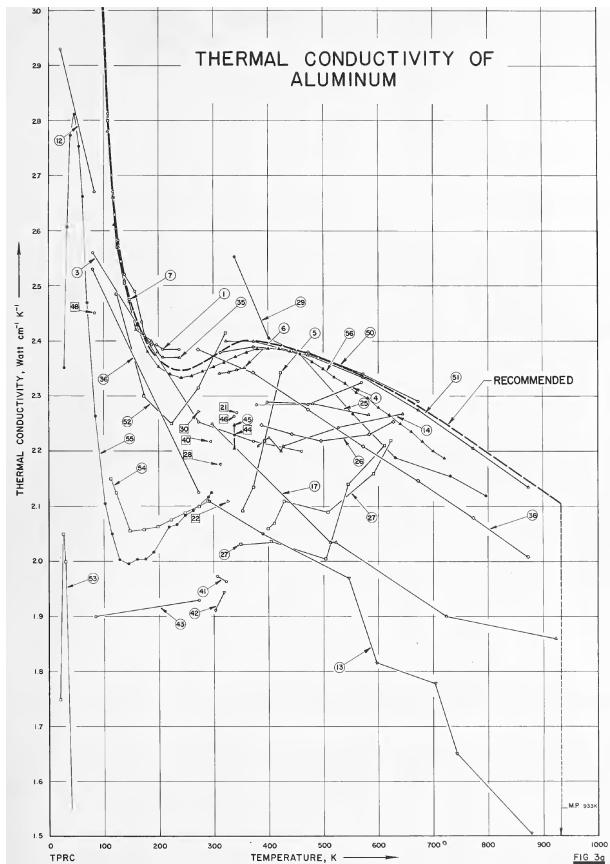


Figure 4.21: Data from [21] on the thermal conductivity of pure aluminum.

of a quadratic fit with unique uncertainties for each data point. While polynomial fits are common in many software packages, it is rare that these can take into account the uncertainty of the data to be fit. For this reason, we recommend that one code fits using the techniques addressed in [23] or similar sources. From that text, the coefficients of the polynomial

$$y = A_1 + A_2x + A_3x^2 \quad (4.14)$$

are found by

$$\begin{aligned} A_1 &= \frac{1}{\Delta} \begin{vmatrix} \sum y_i \frac{1}{u_i^2} & \sum \frac{x_i}{u_i^2} & \sum \frac{x_i^2}{u_i^2} \\ \sum y_i \frac{x_i}{u_i^2} & \sum \frac{x_i^2}{u_i^2} & \sum \frac{x_i^3}{u_i^2} \\ \sum y_i \frac{x_i^2}{u_i^2} & \sum \frac{x_i^3}{u_i^2} & \sum \frac{x_i^4}{u_i^2} \end{vmatrix} \\ A_2 &= \frac{1}{\Delta} \begin{vmatrix} \sum \frac{1}{u_i^2} & \sum y_i \frac{1}{u_i^2} & \sum \frac{x_i^2}{u_i^2} \\ \sum \frac{x_i}{u_i^2} & \sum y_i \frac{x_i}{u_i^2} & \sum \frac{x_i^3}{u_i^2} \\ \sum \frac{x_i^2}{u_i^2} & \sum y_i \frac{x_i^2}{u_i^2} & \sum \frac{x_i^4}{u_i^2} \end{vmatrix} \\ A_3 &= \frac{1}{\Delta} \begin{vmatrix} \sum \frac{1}{u_i^2} & \sum \frac{x_i}{u_i^2} & \sum y_i \frac{1}{u_i^2} \\ \sum \frac{x_i}{u_i^2} & \sum \frac{x_i^2}{u_i^2} & \sum y_i \frac{x_i}{u_i^2} \\ \sum \frac{x_i^2}{u_i^2} & \sum \frac{x_i^3}{u_i^2} & \sum y_i \frac{x_i^2}{u_i^2} \end{vmatrix}, \\ \text{with } \Delta &= \begin{vmatrix} \sum \frac{1}{u_i^2} & \sum \frac{x_i}{u_i^2} & \sum \frac{x_i^2}{u_i^2} \\ \sum \frac{x_i}{u_i^2} & \sum \frac{x_i^2}{u_i^2} & \sum \frac{x_i^3}{u_i^2} \\ \sum \frac{x_i^2}{u_i^2} & \sum \frac{x_i^3}{u_i^2} & \sum \frac{x_i^4}{u_i^2} \end{vmatrix} \end{aligned} \quad (4.15)$$

where each datapoint has a unit total uncertainty  $u_i$ . This expression may be simplified if the uncertainties are uniform. The uncertainties  $A_1$ ,  $A_2$  and  $A_3$  may also be found using the methods shown in [23]. Since this fit involves 3 parameters, a  $3 \times 3$  “error matrix”  $\epsilon$  must be computed. This matrix is found as

$$\boldsymbol{\epsilon} = \boldsymbol{\alpha}^{-1} \quad (4.16)$$

where

$$\alpha_{l,k} = \sum_{i=1}^N \left[ \frac{1}{u_{y_i}^2} x_i^{l-1} x_i^{k-1} \right]. \quad (4.17)$$

Note that this involves the total uncertainty,  $u_{y_i}$ . The diagonal elements of  $\boldsymbol{\epsilon}$  are the standard uncertainties of the fit parameters squared. We note that this method allows for individual uncertainties of each data point. If the uncertainties are unknown and are assumed to be purely random, they may be estimated by

$$r_y^2 = s_y^2 \approx \frac{1}{N-3} \sum_{i=1}^N (y_i - A_1 - A_2 x_i - A_3 x_i^3)^2. \quad (4.18)$$

This can be extended to polynomials of any order.

#### 4.2.8 Uncertainties in Line Fit Parameters

When we limit our attention to the simpler case of a line fit, it becomes possible to write down a closed-form relationship between the input data and the parameter uncertainties. If several measurements are made providing  $N$  values of  $y_i$  as a function of  $x_i$ , and the relationship between them is assumed to be linear,  $y = a + bx$ , then the calibration returns the value of  $a$  and  $b$ .

While many software tools make it simple to fit a line to data, most do not allow one to account for a variable uncertainty for each data point. Bevington and Robinson [23] show that

$$a = \frac{1}{\Delta} \left( \sum \frac{x_i^2}{u_i^2} \sum \frac{y_i}{u_i^2} - \sum \frac{x_i}{u_i^2} \sum \frac{x_i y_i}{u_i^2} \right) \quad (4.19)$$

$$b = \frac{1}{\Delta} \left( \sum \frac{1}{u_i^2} \sum \frac{x_i y_i}{u_i^2} - \sum \frac{x_i}{u_i^2} \sum \frac{y_i}{u_i^2} \right) \quad (4.20)$$

where all sums are over  $i$  from 1 to  $N$  and

$$\Delta = \sum \frac{1}{u_i^2} \sum \frac{x_i^2}{u_i^2} - \left( \sum \frac{x_i}{u_i^2} \right)^2. \quad (4.21)$$

For fixed uncertainty, the fit becomes

$$a = \frac{1}{\Delta'} \left( \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \right) \quad (4.22)$$

$$b = \frac{1}{\Delta'} \left( N \sum x_i y_i - \sum x_i \sum y_i \right) \quad (4.23)$$

where

$$\Delta' = N \sum x_i^2 - \left( \sum x_i \right)^2. \quad (4.24)$$

Bevington and Robinson also provide an analysis based on the Taylor Series Method of the uncertainty of  $a$  and  $b$ . The independent variable,  $x$ , is assumed to be error-free. For a case where all values of  $y_i$  have the same uncertainty  $u_y$ ,

$$\begin{aligned} u_a^2 &= \frac{u_y^2}{\Delta'} \sum x_i^2, \\ u_b^2 &= N \frac{u_y^2}{\Delta'}, \end{aligned} \quad (4.25)$$

If each  $y_i$  value has a unique uncertainty, such as a case where the uncertainty of the measurement is a percentage of the measurement, then

$$u_a^2 = \frac{1}{\Delta} \sum \frac{x_i^2}{u_{y_i}^2}, \quad (4.26)$$

$$u_b^2 = \frac{1}{\Delta} \sum \frac{1}{u_{y_i}^2}. \quad (4.27)$$

Additionally, if a linear fit has been made to data for which the random uncertainty is unknown, the random uncertainty of the dependent variable can be estimated based on the fit [23]. It is necessary to assume that the independent variable is error free and that the uncertainties on the dependent variables are all the same. If a line  $y = a + bx$  has been fit to  $N$  data points, then

$$r_y^2 = s_y^2 \approx \frac{1}{N-2} \sum_{i=1}^N (y_i - a - bx_i)^2. \quad (4.28)$$

The  $N - 2$  is due to the fact that 2 degrees of freedom were used to compute  $a$  and  $b$ .

An example is shown in Fig. 4.18. These are calibration data of a pressure transducer using the manometer in Fig. 4.16 as a calibration source. To perform the calibration, the transducer and the manometer are each plumbed to a vessel in which the pressure can be modified as desired. For many values of pressure, the manometer is read and recorded along with the voltage from the transducer. A table of pairs of pressure and voltage values is made, and Fig. 4.18 shows a plot of this table along with a line fit to the data. The random errors from reading the manometer scale are evident in the departure of the data points from the line.

Equation 4.28 is applied to these data to find that  $s_P$ , or the standard random uncertainty of the manometer, is 0.032 [in Hg]. Our recommendation from Sec. 4.2.1 was to assume that the expanded (95%) uncertainty was 1/2 of the smallest division, or 0.05 [in Hg], meaning that we would expect  $s_P = 0.025$  [in Hg]. This assumption was clearly reasonable.

## 4.3 Dynamic Range

A large dynamic range is desirable for any measurement. Dynamic range is a dimensionless quantity defined as the ratio of the largest measurement possible (Full-scale output, FSO) to the resolution of the measurement. Think of this as the number of possible unique output values of the measurement system. As a simple example, consider a digital outdoor thermometer with a resolution of 1 degree and a readout range from 1 to 100 degrees. The dynamic range of this system is 100.

A capacitive manometer for pressure measurement is an example of an extremely high dynamic range instrument. One particular product claims a resolution of  $1 \times 10^{-6}$  of FSO. Its dynamic range is  $\text{FSO} / 1 \times 10^{-6} = 1 \times 10^6$ .

Many sensors have a fixed minimum uncertainty and FSO. These normally have their accuracy specified as some percentage of FSO. Higher quality instruments have their accuracy stated as a percentage of the measurement. For these, the minimum measurable value is usually stated as a resolution. Again, the minimum resolvable value and the FSO, and thus the dynamic range, of this system is fixed. The user must ensure that the dynamic range of the system is well used by choosing a device with an appropriate FSO for the measurement they are making. For instance, it

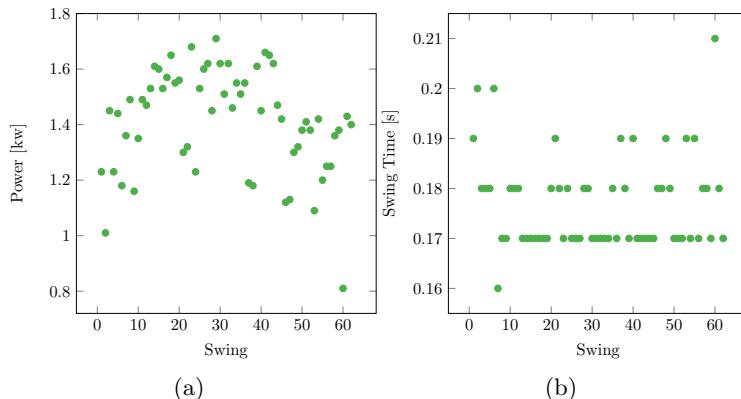


Figure 4.22: Two outputs of a commercial baseball bat sensor: a) Power and b) Time of swing.

would be poor practice to measure pressures in the range of 1 kPa with a sensor having a range of 1 MPa.

Figure 4.22 shows examples of good and bad dynamic range. These are two measurements made by a baseball bat sensor. The first, power, is well resolved. The second, time of swing, has a poor dynamic range relative to the range of values measured. This is easily seen by how many unique values of time appear in the results (only 6).

Some systems allow a user to optimize their dynamic range for a measurement. While the dynamic range of these systems is fixed, the user may adjust the FSO, allowing the user to ensure that they are using all of the dynamic range available. Maximizing the use of the dynamic range of a measurement system minimizes the impact of random uncertainty. We will discuss two examples, the first of which is very common.

The dynamic range of digital data acquisition systems is normally specified by the number of bits  $N_{\text{bits}}$ , and their dynamic range is  $2^{N_{\text{bits}}}$ . For systems with small  $N_{\text{bits}}$ , it is critical to ensure that all of the dynamic range is used.

Most digital data acquisition systems use a successive approximation comparator device consisting in part of a bank of parallel resistors of known value. A reference voltage  $E_{\text{ref}}$  is divided successively by these resistors and the output of this circuit is fed to a digital device called a comparator.

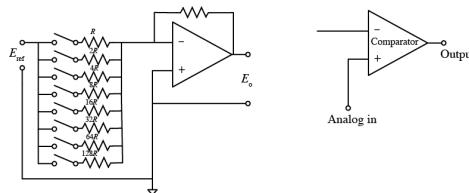


Figure 4.23: Schematic of an analog to digital converter, after [24]. The output of this device connects to a comparator that compares the output of the circuit to the signal being measured.

A comparator outputs low if one input is lower than the other and high when the lower input exceeds the upper input. When this occurs, the bits associated with the open resistors represent zeros while closed resistors represent ones (Fig. 4.23). The system applies a reference voltage  $E_{\text{ref}}$  to the end of the resistor series (which forms a series of voltage dividers) and compares the input voltage to the voltage between each of the resistors.

In many systems,  $E_{\text{ref}}$  may be chosen by the user, effectively allowing them to choose the maximum measurable voltage, or FSO. By choosing the FSO to be larger than the maximum anticipated value, but not larger than necessary, the user may maximize the dynamic range that is used. Failure to do so, especially in cases where the dynamic range is limited, can result in a poorly resolved measurement.

The optical fluid velocity measurement technique, Particle Image Velocimetry (PIV) [25]. This technique senses the displacement of particles suspended in a moving fluid by imaging them at two points in time with a digital camera. The minimum resolvable displacement is a function of many imaging and flow characteristics, but its value may have limits due to constraints that the experimentalist can not improve upon. The dynamic range of the technique then becomes a function of the FSO, which is the maximum displacement of the particles. Therefore, for this measurement, it is critical to maximize the particle displacements.

## 4.4 Effects of Spatial and Temporal Resolution

If a measurement system lacks sufficient resolution, whether that be in time or space, the measured response will be smeared, resulting in uncertainties.

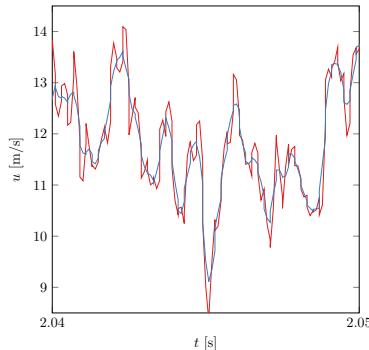


Figure 4.24: Velocity as a function of time measured with a hot wire for a case where (red) the sensor has sufficient temporal resolution and (blue) insufficient temporal resolution.

Unfortunately, if a measurement suffers from resolution issues, there is no way to use that measurement to estimate the impact of the resolution issues, and for this reason, they are often ignored in uncertainty analyses.

An example is shown in Fig. 4.24. The red curve is the unadulterated signal, while the blue curve is affected by a lack of temporal resolution. The difference between the two curves is an error that cannot be addressed through uncertainty analysis since it is a function of the frequency content of the signal. Furthermore, if the blue curve were acquired, there is no way to determine whether the sensor has sufficient temporal resolution. These same issues also affect spatial resolution in a similar manner.

Since the impact of this issue cannot be assessed, it is critical for an experimentalist to ensure that the issue does not exist. The temporal and/or spatial frequency of the signal/field of interest must be known before one can know if the sensor has sufficient resolution.

## 4.5 Total Uncertainty of a Single Variable

For an *a priori* estimate of total uncertainty, ideally, the random uncertainty of the result variable  $X$  may be determined by finding the standard deviation of preliminary measurements. If these are not available, it may be estimated as the resolution of the device based on instrument specifications. For an *a posteriori* analysis, the standard deviation  $s_X = r_X$  should be

used.

The total standard uncertainty is then

$$u_X = \sqrt{r_X^2 + u_{X_1}^2 + u_{X_2}^2 + \dots + u_{X_n}^2}, \quad (4.29)$$

where  $u_{X_1} \dots u_{X_n}$  are the individual  $1 - \sigma$  uncertainties from known sources. Or, if an average of  $X$  is the quantity of interest,

$$u_{\bar{X}} = \sqrt{r_{\bar{X}}^2 + u_{X_1}^2 + u_{X_2}^2 + \dots + u_{X_n}^2}. \quad (4.30)$$

Since the random uncertainty of the mean is root summed with the other uncertainties, it is cost-effective to target a value (through choice of the number of samples) that is similar to the largest of the other uncertainties. There is cost associated with minimizing the uncertainty (i.e. capital cost of equipment) or random uncertainty (i.e. number of samples). Given the root sum nature of the total uncertainty (Eq. 4.30), the return from making either of these quantities significantly smaller than the other will quickly diminish. Once the equipment has been chosen (fixing the  $u_{X_i}$ ) the required number of samples to reduce  $r_{rX}$  to a similar level as  $u_{X_i}$  should be acquired.

This is, of course, a 68% uncertainty. If one desires a 95% uncertainty,  $u_X$  must be expanded using the appropriate value which depends on the assumed PDF of the errors. In [1], it is argued that an overall uncertainty will nearly always constitute a large collection of random variables, and therefore by the Central Limit Theorem, will follow a Gaussian distribution. Therefore,

$$U_{X_{95}} = 1.96u_X, \quad (4.31)$$

and similarly for the mean.

### Example 4.3: Uncertainty of Time-Averaged Pressure Based on a Piezo-Resistive Pressure Sensor

In Sec. 4.2.5, we discussed a pressure sensor with specified a non-linearity uncertainty of 1% of FSO, while non-repeatability and hysteresis were each 0.4% of FSO. As with all instrument specifications, we assume these are 95% confidence values. We are interested in time-averaged pressure, and the actual pressure is known to vary in time. A 2-PSI sensor is used and

preliminary measurement shows the pressure to be 0.1 psi with a sample standard deviation of  $s_P = 0.09$  psi. This makes the random uncertainty of the mean, assuming a large sample of size  $N$ , of

$$s_{\bar{P}} = r_{\bar{P}} = 0.09/\sqrt{N}. \quad (4.32)$$

Uncertainty Source	Spec %FSO	$1 - \sigma$ % FSO	PSI
$r_{\bar{P}}$			$0.090/\sqrt{N}$
Non-linearity	1	0.5	0.010
Hysteresis	0.4	0.2	0.004
Non-repeatability	0.4	0.2	0.004

The root sum of the three known-source uncertainties (traditionally called bias uncertainties) is

$$u_{\bar{P}} = \sqrt{0.010^2 + 0.004^2 + 0.004^2} = 0.011. \quad (4.33)$$

We would like our random of the uncertainty of the mean to have a similar value, so

$$N = \left( \frac{0.09}{0.011} \right)^2 = 61.3 \approx 61 \quad (4.34)$$

Therefore, the total standard uncertainty of the pressure measurement is

$$u_{\bar{P}} = \sqrt{0.011^2 + 0.01^2 + 0.004^2 + 0.004^2} = 0.016. \quad (4.35)$$

Note that these are likely “cheap” samples, acquired with a digital data acquisition system. If this is the case, it may make sense to acquire enough data to render the random uncertainty insignificant. For instance, 10,000 samples would result in  $r_{\bar{P}} = 0.0009$  psi, which is insignificant by the quarter rule. Therefore, the total uncertainty of the pressure measurement is 0.011 PSI.

This value is at the 68% confidence level. We can multiply by 1.96 to find  $U_{\bar{P}_{95}} = 0.0216$  PSI. To express this as a relative uncertainty, we say  $U_{\bar{P}_{95}}/\bar{P} = 0.0216/0.1 = 0.216$  or a 21.6% relative uncertainty. This large value is due to use of a sensor which has a full-scale range much larger than the measurement.

## Summary

You may test your understanding of this chapter by considering the answers to these questions:

- What is the random uncertainty of a measurements whose value remains the same under repeated measurements?
- How could you estimate the uncertainty from known sources using measurements?
- How can one reduce the impact of random uncertainty?
- Is the uncertainty of universal constants zero?
- Is the uncertainty of tabulated material data zero?
- How does the random uncertainty of the mean vary with the number of independent samples?
- What if the samples are not independent?
- Does one need to know the random uncertainty of samples to find the random uncertainty of the mean?

## Homework Problems

1) You have measured a quantity  $X$  five times giving the following result:  
 $X_1 = 10$ ,  $X_2 = 12$ ,  $X_3 = 9$ ,  $X_4 = 1$ ,  $X_5 = 14$ .

If you wish to know the mean of  $X$  to within 2% (95% confidence), based on your initial sample, how many samples are required?

2) What is the dynamic range of a 12-bit camera?

3) What is the dynamic range of a 12-bit DAQ board?

4) What is the dynamic range of a speedometer that goes to 120 mph in 5 mph increments?

5) You make the five measurements below:

100

120

110

70

110

If you want to determine the mean of this variable to within 2% at 68% confidence, what is your best estimate for the number of samples that would be required?

6) You are making pressure measurements with a sensor that has specified uncertainty of 0.5% of reading. For a given test condition, the readout of the sensor says 2000 psi, and based on the variation in the reading, you estimate that your random uncertainty is  $r_P = 40$  psi. If you plan on acquiring the output of this sensor digitally, how many independent samples should you acquire to make the best use of your instrument and your time?

7) What is your voltage resolution if your voltage range is  $\pm 5V$  and the A-D converter has 8 bits?

8) You have a bourdon-tube pressure meter, which is NIST traceable. You wish to calibrate a pressure transducer against it. This pressure transducer is the diaphragm type, with a full bridge strain-gage arrangement. The excitation voltage is  $\pm 10$  V and an amplifier of gain 10 is used. You have taken the 21 data points shown in the table below in the order in which they are listed. Based on this data, discuss the severity of the following error sources:

- a) Hysteresis
- b) Noise
- c) Zero Error
- d) Sensitivity Error

P[psi]	Volts
0.0000	0.10000
2.0000	2.1100
4.0000	4.1000
6.0000	6.3500
8.0000	8.3500
10.000	10.300
8.0000	7.9500
6.0000	5.7200
4.0000	3.7500
2.0000	1.7000
0.0000	-0.02000
2.0000	2.1000
4.0000	4.2000
6.0000	6.3000
8.0000	8.4000
10.000	10.200
8.0000	7.9000
6.0000	5.8000
4.0000	3.7000
2.0000	1.7000
0.0000	-0.0100000

- 9) If you have an uncertainty  $u_{X_1}$  from one error source and another  $u_{X_2}$  which is 4 times larger, what percentage of the overall uncertainty of  $X$  comes from  $u_{X_1}$ ? (Hint: compare the overall uncertainty with and without  $u_{X_1}$ )
- 10) If you have three error sources resulting in uncertainties  $u_{X_1} = u_{X_2}$  from one error source and another  $u_{X_3}$  which is 4 times larger, what percentage of the overall uncertainty of  $X$  comes from  $u_{X_1}$ ?
- 11) A young baseball player uses a bat sensor to measure the speed of the tip of the bat for all their swings. The file BatSpeedSS.txt contains 15 readings of the sensor. Find the player's average tip speed and the uncertainty of that speed to 95% confidence.  
 Repeat the exercise with the file BatSpeed.txt which contains a larger sample.
- 12) Explore the impact of errors on line fits by writing a code that
- Generates  $N$  data points that are evenly spaced along  $x$  from 1 to 8 and follow  $y = 1 + 2x$ . Add errors to the  $y$  values that are normally

distributed around 0 and an arbitrary standard deviation  $s$  that may be fixed or a percentage of  $y$ .

2. Fits a line to the points.
3. Computes the uncertainty of the slope and intercept of the line.
4. Estimates the average uncertainty in  $y$  using Eq. 4.28.

Try this with different values of  $s$  and  $N$ , and draw some conclusions about how errors in the data points impact the uncertainty of the parameters. Consider the validity of Eq. 4.28. Note that the part of the code that generates the data has no connection to the part that analyzes the data.

## Chapter 5

# Propagation of Multivariable Uncertainties

Often, experiments involve more than one measurement that are combined into a final result. Perhaps the simplest example is determining the lost mass during a process by weighing the sample before and after and computing the difference,

$$m_{\text{lost}} = m_2 - m_1. \quad (5.1)$$

Similarly, the velocity of a moving fluid stream can be determined using a Pitot-static probe and measuring the difference in pressure between the total and static port,

$$V = \sqrt{\frac{P_t - P_s}{2\rho}}, \quad (5.2)$$

where  $\rho$  is the fluid density.

In any case where the final result is a function of multiple variables that are subject to error, the uncertainty of each variable must be propagated into the result. This can be done using two methods, each with advantages and disadvantages. The first is the Taylor Series Method (TSM), and the second is the Monte Carlo Method (MCM).

To use either of these, one must know the relationship between the variables. For instance, if the outcome variable  $z$  is a function of  $x$  and  $y$ , that exact relationship must be known. The relationship between  $z, x$ , and

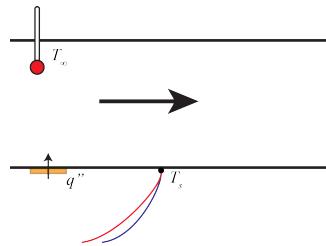


Figure 5.1: Experiment to determine the heat transfer coefficient  $h$  from Newton's law of cooling, [5.3].

$y$  is called the Data Reduction Equation (DRE) and Equations 5.1 and 5.2 are examples. While this may sound trivial, in our experience, failure at this stage of an uncertainty analysis is common for two reasons:

1. An incorrect DRE is assumed
2. The DRE is complex and unknown

An example of the first reason is failing to account for all physics in a process. If one is trying to measure the convection heat transfer coefficient, as shown in Fig. 5.1, Newton's Law of Cooling can be used [26],

$$q'' = h(T_s - T_\infty), \quad (5.3)$$

where  $q''$  is heat flux,  $h$  is the convection heat transfer coefficient,  $T_s$  is the surface temperature and  $T_\infty$  is the free stream fluid temperature. This assumes, perhaps incorrectly, that all heat leaving the surface is due to fluid convection. If radiation heat transfer is also occurring to the surroundings at  $T_\infty$ ,

$$q'' = h(T_s - T_\infty) + \epsilon\sigma(T_s^4 - T_\infty^4), \quad (5.4)$$

where  $\epsilon$  represents the surface emissivity and  $\sigma$  is the Stephan Boltzmann constant. If the second term is significant, but Eq. 5.3 is used for the DRE, the uncertainty analysis will be incorrect.

## 5.1 Taylor Series Method of Uncertainty Propagation

This method uses a Taylor series expansion about small variations in each of the input variables to determine the uncertainty of the output variable [1]. Again, assuming  $z = f(x, y)$

$$u_z^2 = u_x^2 \left( \frac{\partial z}{\partial x} \right)^2 + u_y^2 \left( \frac{\partial z}{\partial y} \right)^2 + 2u_{xy} \left( \frac{\partial z}{\partial x} \right) \left( \frac{\partial z}{\partial y} \right). \quad (5.5)$$

It is important to ensure that all uncertainties are at the same confidence level. While there are several valid approaches, the default approach used here will be to propagate standard (68% coverage) uncertainties.

The derivatives form the sensitivity of the result to each variable. The final term contains the covariance of  $x$  and  $y$ ,  $u_{xy}$ . Likely due to difficulties in estimating its value, this term is often ignored. We believe that it is often important and should always be considered. Note that, unlike the first two terms, the covariance term can be negative or positive depending on the signs of the derivatives. Although rare, it is also possible for the covariance itself to be negative. This means that correlation between error sources can increase or decrease the uncertainty.

More generally, the experimental result can be a function of any number of variables. Let us assume that  $Y = f(X_1, X_2, \dots, X_j)$  where the subscripts refer to individual measured variables rather than sample numbers. We refer to the derivation in the appendix of [2] and also adopt their nomenclature that  $\partial Y / \partial X_i = \theta_{X_i}$ . While we will not repeat the derivation, it is important to note the assumptions that were made, which are that the errors in each variable are small and that the derivatives are of “reasonable size” so that the higher order terms of the series become negligible.

The uncertainty of  $Y$  is

$$u_Y^2 = \sum_{i=1}^j \theta_{X_i}^2 u_{X_i}^2 + 2 \sum_{i=1}^{j-1} \sum_{k=i+1}^j \theta_{X_i} \theta_{X_k} u_{X_i X_k} \quad (5.6)$$

The terms  $u_{X_i X_k}$  are problematic yet important. These are the covariances of  $X_i$  and  $X_k$  and account for the extent to which errors in  $X_i$  and  $X_k$  are correlated. The definition of the covariance is

$$u_{X_i X_k} = \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{j=1}^N \delta_{X_{ij}} \delta_{X_{kj}} \right), \quad (5.7)$$

where  $N$  is the number of samples over which the covariance is computed and  $\delta_{X_{ij}}$  is the instantaneous error of  $X_i$ . In most cases, these errors are not known. Random fluctuations contributing to the random uncertainty can also be correlated between different variables in rare cases. One strategy for handling these is to compute the random uncertainty of the final result ( $z$  in Eq. 5.10) as suggested in [2].

The most obvious way that two measurements can have correlated error sources is when the same instrument, or parts of a measuring system, are used for more than one measurement. In Eq. 5.1, if the sample is weighed on the same mass scale before and after the process, there is correlated error between the two measurements. Clearly, errors from the calibration of the scale will impact both measurements similarly.

For camera-based measurements, the same region of the camera sensor may affect measurements at more than one location. In this case, measurements at these locations will have some level of correlation.

Coleman and Steele [1, 2] advise that correlations are treated as perfect (meaning  $u_{X_i X_k} = u_{X_i} u_{X_k}$ ) if the same instrument is used for both measurements and zero otherwise. This is a vast improvement over ignoring correlations and works well for cases where instruments are used for multiple measurements. A more subtle example is two sensors that each have an error that stems from a change in the temperature of the sensor. In this case, a change in the ambient temperature will cause an error in both measurements simultaneously. In this case, the correlation could be positive or negative.

It is possible that two measurements each have multiple error sources. In such cases, some error sources may be correlated to one another while others are not. Say that we are to measure  $X$  and  $Y$  and we denote the systematic uncertainties of these from error source 1 as  $u_{X,1}$  and  $u_{Y,1}$ , The systematic uncertainties of these from error source 2 are  $u_{X,2}$  and  $u_{Y,2}$ , and so on. In this case, if all error sources of  $X$  and  $Y$  are correlated,

$$u_{X,Y} = \sum_{\alpha=1}^L (u_{X,\alpha})(u_{Y,\alpha}), \quad (5.8)$$

where  $L$  is the total number of correlated error sources [1]. If an error source is not common to the measurement of  $X$  and  $Y$ , it should not be included in Eq. 5.8.

### Example 5.1: Measuring the heat transfer coefficient for a heated wall in a wind tunnel

We return to our example of Eq. 5.3 as the basis for a DRE. We will begin with a simple case and gradually increase the level of complexity. The first step is to find the DRE. Equation 5.3 must be solved for the quantity we will later determine experimentally, in this case,  $h$ .

$$h = \frac{q''}{T_s - T_\infty} \quad (5.9)$$

Consider an experiment where the wall temperature is measured by a thermocouple built into the wall, and the free stream temperature is measured with a traditional analog bulb thermometer. The heat flux is measured with a heat flux sensor built into the wall. With all three measurements being completely independent of one another (no common equipment or calibrations), it is unlikely that their uncertainties are correlated. Eq. 5.10 becomes

$$u_h^2 = \theta_{q''}^2 u_{q''}^2 + \theta_{T_s}^2 u_{T_s}^2 + \theta_{T_\infty}^2 u_{T_\infty}^2. \quad (5.10)$$

The three derivatives are  $\theta_{q''} = 1/(T_s - T_\infty)$ ,  $\theta_{T_s} = -q''/(T_s - T_\infty)^2$ , and  $\theta_{T_\infty} = q''/(T_s - T_\infty)^2$ . Since all of these derivatives will be squared in the TSM, their signs are of no consequence (we will see that this is not true when the correlation terms are included).

Note that the uncertainty of  $h$  is a function of the nominal values of the variables through the derivatives. As such, a nominal measurement is required for the uncertainty analysis. Additionally, random uncertainties must be computed from acquired data. Two of these quantities,  $T_\infty$  and  $q''$  do not vary significantly in time and therefore have no random uncertainty. A preliminary measurement is made resulting in the following values:  $T_\infty = 20^\circ\text{C}$ ,  $T_s = 70^\circ\text{C}$ , and  $q'' = 500\text{W}$ . Given these values, the heat transfer coefficient  $h = 10 \text{ W/m}^2\text{K}$ .

Now, let us consider the uncertainty components of each of these quantities. The heat flux sensor has a dedicated acquisition unit and the

Table 5.1: Measured values, uncertainties and sensitivities of the three variables

Term	Values	$u$	$\theta$	$\theta^2 u^2$
$q''$	500W	3.75W	$0.02 \text{ } ^\circ\text{C}^{-1}$	$0.0056 \text{ W/m}^2$
$T_\infty$	$20^\circ\text{C}$	$0.1^\circ\text{C}$	$0.2 \text{ W/m}^2/\text{ }^\circ\text{C}^2$	$4 \times 10^{-4} \text{ W/m}^2$
$\bar{T}_s$	$70^\circ\text{C}$	$0.6^\circ\text{C}$	$-0.2 \text{ W/m}^2/\text{ }^\circ\text{C}^2$	$0.014 \text{ W/m}^2$

user reads the measured value from a display. Its value does not vary during the experiment, so  $r_{q''} = 0$ . The device's specification sheet claims an accuracy of 1.5% of reading, so  $2u_{q''} = 0.015q''$ .

The bulb thermometer used to measure  $T_\infty$  has an “accuracy” of  $0.1^\circ\text{C}$  and has divisions twice that far apart. The temperature value is read off an analog scale by the experimentalist. It is not reread for every sample from the  $T_s$  or  $q''$  data acquisition system. The common guidance for this circumstance is that a standard total uncertainty equal to half of the smallest division should be assigned [1].

The wall temperature is measured by a high-speed data acquisition unit coupled to an analog preprocessing circuit. The specifications list many uncertainty specifications, including that the cold junction compensation is accurate to  $\pm 0.6^\circ\text{C}$  (typical) or  $\pm 1.2^\circ\text{C}$  max. Clearly, neither of these numbers represent a standard deviation. While it is difficult to be sure, these numbers seem to represent a 68% and 95% coverage values. This uncertainty will dominate all others for this measurement by the quarter rule. The temperature is averaged over 1000 independent samples. The standard deviation of these measurements is  $2^\circ\text{C}$ . The random uncertainty of the mean wall temperature is thus  $r_{\bar{T}_s} = s_{\bar{T}_s} = s_{T_s}/\sqrt{1000} = 0.063^\circ\text{C}$ . This is also insignificant by the quarter rule, so  $u_{\bar{T}_s} = 0.6^\circ\text{C}$ .

The nominal measurements, their uncertainties, their derivatives and the derivative times the sum of the squared uncertainties are shown in Table 5.1. The square root of the sum of the right most column gives us a total uncertainty of  $0.143 \text{ W/m}^2\text{K}$ . The table makes it clear that the uncertainty of the bulb thermometer does not contribute, while the uncertainties from both the thermocouple and the heat flux sensor are significant.

It is suggested that the bulb thermometer be replaced by a thermocouple probe which is to be read by the same analog preprocessing unit and data acquisition system as the wall temperature. As a result, these two measurements will now have a correlated uncertainty. In other words, whatever the unknown fixed error is due to the cold junction compensation, it will be the same for both thermocouple measurements. In this situation,  $u_{T_s T_\infty} = u_{T_s} u_{T_\infty} = u_{T_s}^2$ . The free stream temperature is now measured repeatedly but the random uncertainty of the temperature measurement has already been shown to be negligible. The random uncertainty of the mean of  $T_\infty$  is as small as that of  $T_s$ . The TSM equation, with the terms with zero random uncertainty dropped, becomes

$$\begin{aligned} u_h^2 &= \theta_{q''}^2 u_{q''}^2 + \theta_{T_s}^2 u_{T_s}^2 + \theta_{T_\infty}^2 u_{T_\infty}^2 + 2\theta_{T_s} \theta_{T_\infty} u_{T_s T_\infty}, \\ &= \theta_{q''}^2 u_{q''}^2 + 2\theta_{T_s}^2 u_{T_s}^2 - 2\theta_{T_s}^2 u_{T_s}^2, \\ &= \theta_{q''}^2 u_{q''}^2. \end{aligned} \quad (5.11)$$

where in the second line, we have taken advantage of the fact that  $b_{T_s} = b_{T_\infty}$ ,  $\theta_{T_s} = -\theta_{T_\infty}$ . The correlation between the two temperature measurements and the fact that their derivatives have opposite signs means that their systematic uncertainties no longer contribute to the overall uncertainty. Taking this correlation into account,  $u_h = 0.0748 \text{ W/m}^2\text{K}$ .

The surprising result is that we are able to reduce the uncertainty by replacing a more accurate device with a less accurate one since the replacement device has a systematic uncertainty correlated to another device. Our ability to do so relied on a DRE that produced oppositely signed derivatives of the correlated measurements. Additionally, if the random uncertainties, which are less likely to be correlated, were large, we would likely not have seen a benefit. For a DRE where pairs of measurements have positive derivatives, correlation will cause the overall uncertainty to be larger.

Clearly, this analysis revealed benefits from correlated uncertainties. This is not uncommon, and for this reason, we recommend considering correlation at all stages of analysis, whether it be planning, debugging, or determining the final experimental uncertainty.

**Example 5.2: Measuring volume flow rate based on point velocity measurements**

The measurement of volume flow rate by averaging of velocity measurements made at various points in space presents an interesting example of the effects of correlated uncertainties. Volume flow rate is the spatial average of velocity, or

$$Q = A \sum_{i=1}^N V_i, \quad (5.12)$$

where  $A$  is the area of the plane encompassing the velocity measurements  $V_i$ . We will assume no uncertainty in  $A$  and consider the possibilities of perfect correlation in the  $V_i$  values as well as no correlation. Assuming that all quantities are sufficiently averaged in time to make random uncertainties negligible, and applying Eq. 5.6,

$$u_Q^2 = A^2 \left[ \sum_{i=1}^N \left( \frac{1}{N} \right)^2 u_{V_i}^2 + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N \frac{1}{N} \frac{1}{N} u_{V_i} u_{V_k} \right]. \quad (5.13)$$

If the uncertainties of the velocities are equal and there is no correlation between the velocity measurements, the last term is zero and

$$u_Q = A \frac{u_V}{\sqrt{N}}, \quad (5.14)$$

which says that the uncertainty in the volume flow rate decreases with the number of velocity measurements. This result may seem intuitive, but consider what is required for velocity errors to be uncorrelated. The various  $V_i$  measurements would need to each be made with different probes connected to different sensors. Correlated velocity measurements are much more common.

In this case where all velocity uncertainties are correlated, again assuming the uncertainties are the same for each measurement,

$$u_Q = A u_V. \quad (5.15)$$

This says that the flow rate uncertainty is the same as the velocity uncertainty scaled by area and that more data points will not improve the result.

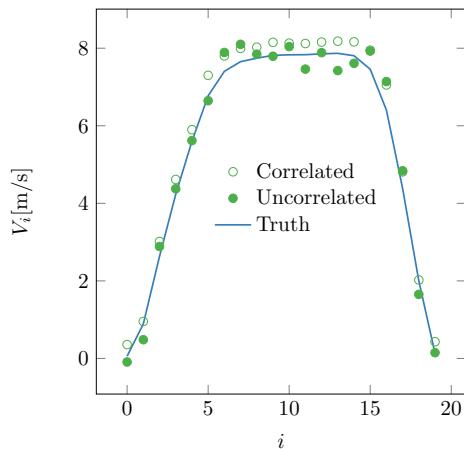


Figure 5.2: Velocity profile of a jet comparing a case where the errors between the individual velocity measurements are correlated or uncorrelated compared to the true velocities.

This is because the derivatives in Eq. 5.13 are all positive and correlation makes uncertainty larger.

The difference between these two cases is illustrated in Fig. 5.2. For errors that are correlated, all velocity measurements have the same error. Thus, when the velocity is numerically integrated over the full profile, the error in volume flow rate scales with the error in velocity. When the errors are uncorrelated, the integration process will mitigate the errors and more points in the profile will improve the result.

In Coleman and Steele's text [1], several examples are given for which an experiment that is assumed steady is actually unsteady. In such a situation, multiple variables may change in a correlated manner. In the previous example, if the heat flux is fixed but the velocity over the plate increases,  $h$  will increase, and  $T_s$  will decrease as a result. These authors recommend against the propagation of uncertainties in such a situation since correlations of drifting variables are difficult to analyze correctly.

In some cases, the DRE may not be available analytically. For instance, the DRE could be embodied in proprietary software. In such cases, TSM is still possible if one can list the variables on which the result depends and if the uncertainties of these variables are known. In this case, the

derivatives may be estimated through finite differences. One should be careful to ensure that the derivatives are evaluated for the nominal input variable values of interest.

One major limitation of the TSM is that all input uncertainties are assumed to be symmetric and the output distribution can only be symmetric. This is often of no concern, but in other cases, can become a major limitation. Consider the application of the TSM to the standard deviation operator, Eq. 3.14, which is repeated here,

$$s_X = \left[ \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \right]^{1/2}.$$

We could compute the derivatives of  $s_X$  with respect to each of the samples in a similar manner that we did with the mean in Chapter 4 and find the uncertainty of the standard deviation. The TSM would return a symmetric interval inside of which the true uncertainty of the standard deviation should lie 68% of the time. However, it is not possible for errors in the standard deviation to be negative due to errors in the samples as a result of the deviations from the mean being squared. There is no central tendency in this case, and a symmetric uncertainty is not appropriate.

## 5.2 Monte Carlo Method of Uncertainty Propagation

The Monte Carlo Method (MCM) of uncertainty propagation is a numerical technique that has many benefits and a few drawbacks compared to the TSM. While numerical, it is simple enough to be performed on a spreadsheet.

The MCM has the side benefit of helping one visualize how measurement uncertainty is related to error distribution. Each individual uncertainty is used to generate simulated errors on that measurement from a distribution of appropriate shape and width.

In the simplest case, a DRE is known and an estimate of the uncertainty of each input measurement is known. In this case, the MCM procedure is as follows:

1. A random number is drawn from the assumed error distribution of each measurement. In most cases, this random draw is from a

Gaussian distribution with a standard deviation corresponding to the standard uncertainty of that measurement.

2. These simulated random errors are added to the nominal value for each measurement and the error plus nominal value is used to compute the output quantity through the DRE.
3. This is repeated thousands of times to generate a distribution of the output quantity.
4. If the output distribution is judged to be Gaussian, as will often be the case, the uncertainty can be expressed as the standard deviation of the results.
5. For cases with a highly nonlinear DRE, the output distribution may not be Gaussian. In such cases, the range in which 95% of the values are found must be determined.

This procedure is presented visually in Fig. 5.3. In this example, we perform an MCM simulation for an experiment in which four quantities are measured:  $X_1, X_2, X_3$  and  $X_4$ . The quantities  $X_1, X_2$ , and  $X_4$  are assumed to have errors that follow a Gaussian distribution, while the errors on  $X_3$  are assumed to follow a uniform distribution. The width of these distributions is the uncertainty of that variable. Each iteration of the simulation is indicated with the second subscript. Random numbers representing errors are drawn from the assumed distribution for each of the input variables and added to the nominal value for that variable, and the DRE is used to compute  $Y$  for that iteration based on the  $X$  values. The output of an MCM simulation is the distribution of  $Y$ , which will be analyzed as described below to determine an uncertainty.

To demonstrate, we will first repeat the convection example and use Eq. 5.9 as the DRE. The nominal and uncertainty values are found in Table 5.1.

We assume that all errors have zero mean. We will denote a random draw from a Gaussian distribution with zero mean and a standard deviation of  $u$  as  $\mathcal{N}(u)$ . In this case, the random uncertainties are insignificant, so

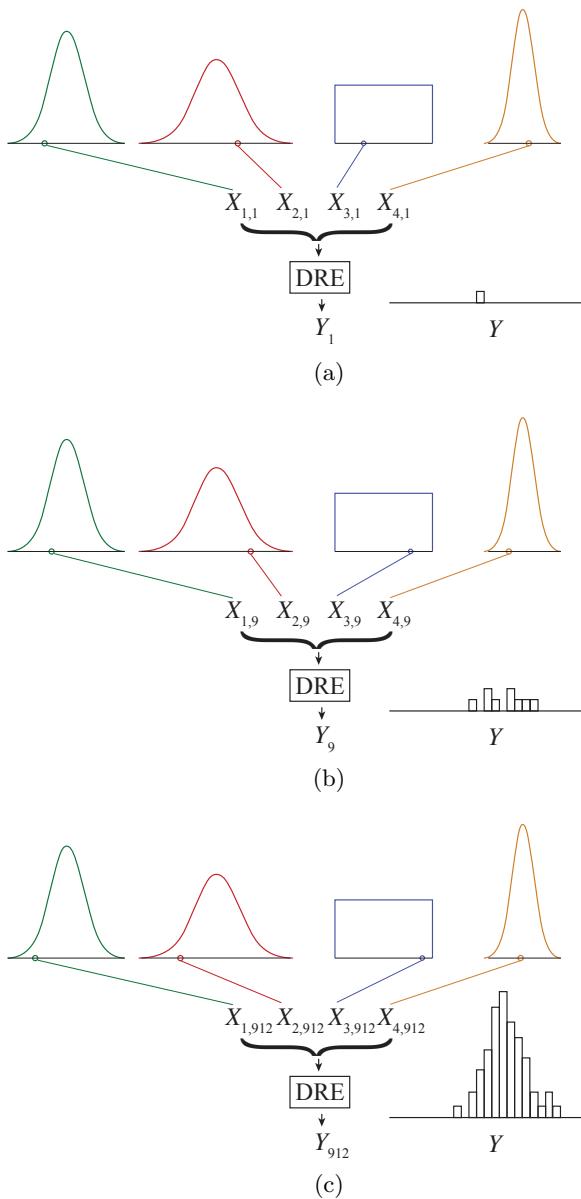


Figure 5.3: The first, 9th and 912th iteration of an MCM analysis based upon a DRE with 4 variables.

for each variable,  $u \approx b$ . Therefore, for each iteration  $i$ ,

$$\begin{aligned} h_i &= \frac{q'' + \mathcal{N}(u_{q''})}{T_s + \mathcal{N}(u_{T_s}) - T_\infty - \mathcal{N}(u_{T_\infty})} \\ &= \frac{500 + \mathcal{N}(7.5)}{70 + \mathcal{N}(1) - 20 - \mathcal{N}(0.14)}. \end{aligned} \quad (5.16)$$

This operation was repeated 10,000 times using the MATLAB script `ConvectionMCM.m`. The first several iterations are shown in Fig. 5.4. Note that each iteration is the nominal value plus a random fluctuation (error). The  $h$  values are computed from Eq. 5.9 using the nominal inputs plus their errors.

Iteration	$T_\infty$	$T_s$	$q''$	$h$
1	19.8868	69.8417	497.2601	9.9542
2	19.9696	70.5518	500.1277	9.8874
3	19.8903	69.8713	507.3031	10.1499
4	20.0492	70.1683	500.1422	9.9791
5	20.0895	70.1504	505.4646	10.097
6	20.0657	69.7727	498.647	10.0317
7	19.8658	69.4986	499.3486	10.0609
8	20.0559	70.346	501.4169	9.9705
9	20.0797	70.6495	492.9962	9.7488
10	20.0539	70.0991	503.6893	10.0647
11	19.921	69.5324	498.6666	10.0515
12	19.9797	70.0888	502.9123	10.0364
13	20.0082	69.7558	498.8508	10.0276

Figure 5.4: The first 13 iterations of Eq. 5.16.

All 10,000 iterations are plotted in Fig. 5.5. The standard deviation of these simulated values of  $h$ ,  $u_h = 0.1425$ , which is very close to the 0.143 value found by TSM. At 10,000 iterations, the answer may vary slightly from one simulation to the next, and, of course, using more iteration will limit this variation.

This simulation assumed that there is no relationship between the variables, or that they are uncorrelated. If we wish to simulate correlation between the two temperature measurements, we simply use the same random draw of temperature error for both temperatures. The first 13 iterations of such a simulation are shown in Fig. 5.6. Note that for each iteration, the variation for both temperatures is in the same direction. This correlation leads to a substantially more narrow distribution (see Fig. 5.7) and the uncertainty is reduced to  $u_h = 0.074$ , very close to the value of 0.0748 from TSM.

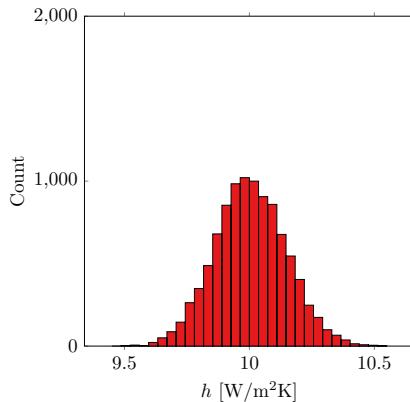


Figure 5.5: Output distribution from MCM analysis of Eq. 5.16 using uncertainties from Table 5.1.

Iteration	$T_\infty$	$T_s$	$q''$	$h$
1	19.9661	69.9661	499.0613	9.9812
2	19.9475	69.9475	496.1081	9.9222
3	19.9464	69.9464	499.3413	9.9868
4	20.0368	70.0368	497.9034	9.9581
5	19.9748	69.9748	498.0904	9.9618
6	20.0872	70.0872	499.5441	9.9909
7	19.9526	69.9526	499.1383	9.9828
8	20.0276	70.0276	501.3358	10.0267
9	20.0737	70.0737	499.9655	9.9993
10	20.0526	70.0526	498.8288	9.9766
11	19.9688	69.9688	501.816	10.0363
12	19.8953	69.8953	501.0253	10.0205
13	20.0161	70.0161	501.9403	10.0388

Figure 5.6: The first 13 iterations of Eq. 5.16 with the errors on  $T_\infty$  and  $T_s$  correlated.

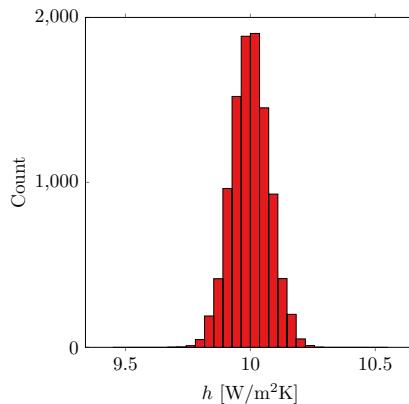


Figure 5.7: Output distribution from MCM analysis of Eq. 5.16 using uncertainties from Table 5.1 and assuming that errors on the two temperatures are correlated.

The more iterations that are performed, the more accurate the result will be. MCMs are inexpensive and it is not difficult to perform enough iterations, typically many thousands, to ensure a converged result. We caution that this may not be practical on a spreadsheet, although spreadsheets with a smaller number of iterations are an excellent way to visualize the process.

The resultant distribution of an MCM will be wider than one would see in an actual experiment. This is because, by their nature, most uncertainties stem from errors that are not resampled during an experiment, so one does not see their distribution. Hypothetically, if all the sources of uncertainty were replaced many times during an experiment, the distribution of experimental results would resemble the MCM result.

For the convection DRE, we have not taken advantage of the main benefits of MCM: the ability to prescribe a non-Gaussian error distribution, the elimination of the need to compute derivatives (which are simple in this case) and the ability to allow an asymmetric uncertainty (which this case does not produce). Each of these advantages are better demonstrated in the case of a more complex and non-linear DRE.

Consider MCM as applied to Eq. 5.4. In addition to being a non-linear

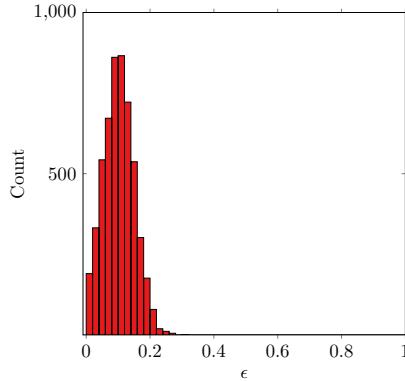


Figure 5.8: Histogram of simulated errors in  $\epsilon$ .

equation for  $h$  in terms of  $T_s$ , this equation also involves the bounded variable for emissivity. By definition,  $0 < \epsilon < 1$  [26]. If  $\epsilon = 0.1$  and  $r_\epsilon = 0.05$ , drawing simulated errors from a Gaussian distribution will result in non-physical negative values. It is a simple matter to modify the Gaussian distribution to place bounds on the results in an MCM analysis. This is shown in Fig. 5.8.

As another example of a non-linear DRE, we turn to baseball bat certification. In the USA, baseball is traditionally played with bats made of wood. Modern materials such as aluminum and/or composites allow bats to be built with less connection between their weight and length. In addition, such bats can have a much larger coefficient of restitution (*COR*) than wooden bats. In order to protect defensive players from hard-hit balls at short distances, bat standards have been enacted in many baseball leagues that limit the performance of bats. One such standard, called the Bat-Ball Coefficient of Restitution (BBCOR) standard, is enforced by testing the *BBCOR* of the bat impacted with a baseball at game speed:

$$\begin{aligned} BBCOR &= \frac{v_i + v_r}{v_i} \left( \frac{m}{M_e} + 1 \right) - 1 \\ M_e &= \left[ \frac{1}{W} + \frac{(Q + 6 - BP)^2}{[I - W(BP - 6)^2]} \right]^{-1}, \end{aligned} \tag{5.17}$$

Table 5.2: Measured values and uncertainties of the variables.

Variable	Values	$u$	$r$
$v_{in}$	150 mph	0	0.5 mph
$v_{out}$	30 mph	0	0.5 mph
$m$	5.3 Oz	0.05 oz	0
$W$	34 Oz	0.05 oz	0
$Q$	22 in	0.125 in	0
$BP$	20 in	0.125 in	0
$MOI$	10,000 Oz in <sup>2</sup>	3.75% of reading	0

where  $I$  is the bat moment of inertia,  $M_e$  is the effective mass of the bat,  $W$  is the mass of the bat,  $BP$  is the balance point of the bat,  $v_i$  is the inbound ball speed,  $v_r$  is the rebound ball speed,  $m$  is the mass of the ball, and  $Q$  is the distance from the impact location and pivot point location.

There are seven quantities to be measured. Nominal values for each and assumed uncertainties are shown in Table 5.2. Note that while the nominal values are representative of real baseball bats, our assumed uncertainties may not reflect values used in real bat testing. Also note that for this example, the random and other uncertainties are listed separately. A total uncertainty must be formed for each quantity before the MCM is performed.

While the other variables are assumed to have errors from a Gaussian distribution, errors in the velocities of the incoming and outgoing ball will be treated as a uniform distribution. For instance, if the velocity measuring instrument has a resolution of 1 mph, the error from digitizing the result is anywhere from 0 to 1 mph and all values in between are equally likely. It is not possible to mix different PDFs using TSM, but for MCM, this presents no problem. We simply draw errors from a uniform distribution rather than a Gaussian. The histograms of each variable and the resultant  $BBCOR$  distribution are shown in Fig. 5.9 a-h.

For these nominal values,  $BBCOR = 0.5091$ . The resulting 10,000 values of  $BBCOR$  are first sorted and the 250th and 9750th values are found to be 0.4886 and 0.5502. These are indicated by the vertical lines in Fig. 5.9h. This illustrates one of the primary advantages of MCM

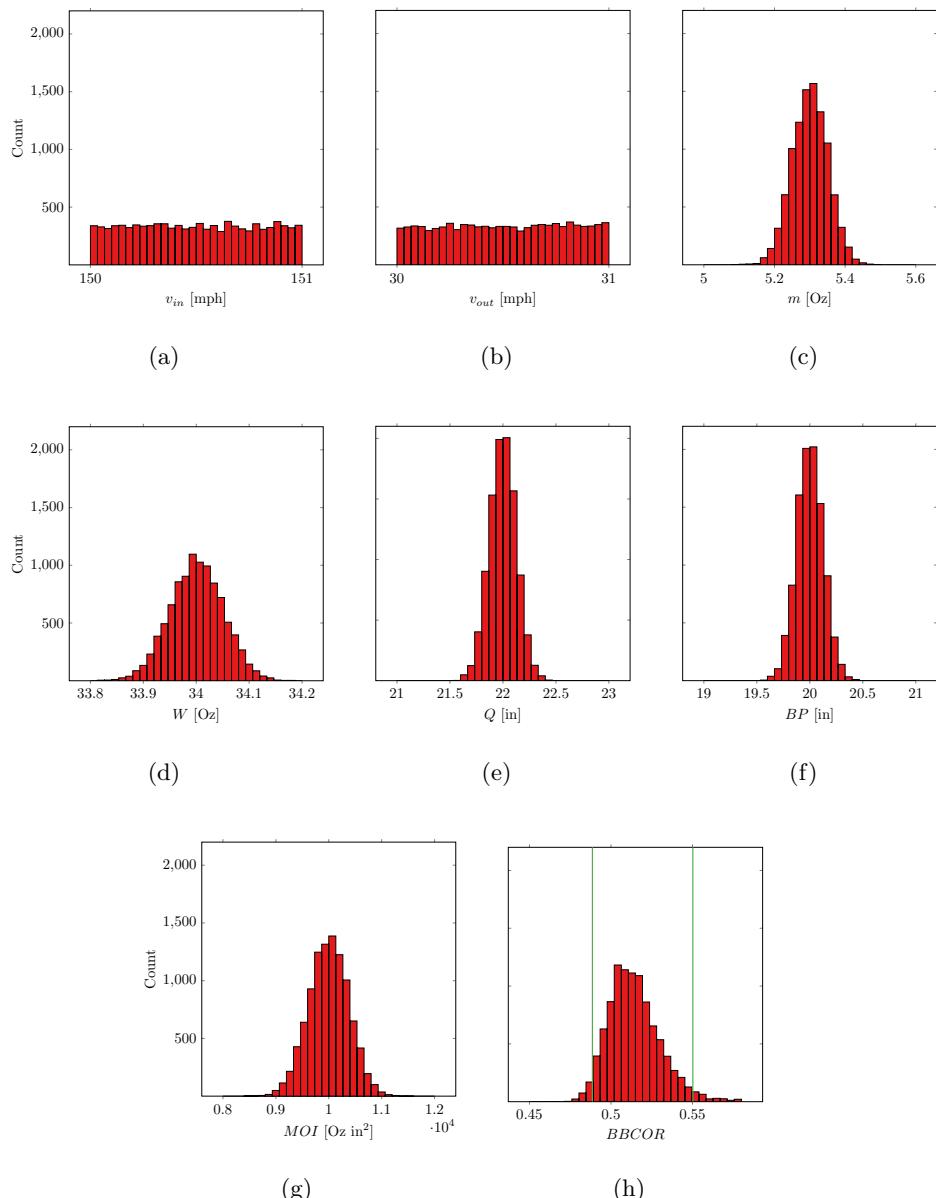


Figure 5.9: Histograms of the seven input quantities (a-g) as well as the simulated output error (h). The green lines indicate the range of 2.5% and 97.5% of the results.

- that the output distribution may be skewed and a unique upper and lower uncertainty band can be found. This is especially important if the uncertainties are large or if the DRE is highly non-linear. In other words, for cases where the assumptions used to form the TSM are violated.

Several pairs of measurements may have error sources that are correlated. As stated above, the primary circumstance that generates correlation is the use of the same instrument for multiple measurements. Looking at the DRE (Eq. 5.17), several masses, lengths as well as a pair of velocities are measured. If, for instance, the same scale is used for multiple mass measurements, the errors between these measurements are correlated. Since two mass measurements appear as a ratio, the derivative of *BBCOR* with respect to two masses has opposite signs, so correlation is beneficial.

To simulate correlation in an MCM analysis, one simply uses the same random draw for more than one variable to simulate perfect correlation. If a less than perfect correlation is desired, the common random draw may be weighted and added to another, independent random draw.

Now we pose this question: If this experiment were performed, would the distribution of *BBCOR* values measured appear like Fig. 5.9h? The answer is “no,” since this MCM is, in the parlance of Moffat [9], an *Nth* order analysis and simulates all instruments being replaced for each iteration. To simulate what one would measure, only random uncertainties should be entered into the simulation, since only those error sources are resampled during the experiment.

### Example 5.3: Measurement of Baseball Bat Tip Speed

A baseball player uses a bat sensor to measure the speed of the tip of the bat  $V$  for all their swings. The average of several swings is 53 mph. The sensor resides on the knob of the bat and directly measures the rotation rate around the pivot point or, roughly, the point where the hands grasp the bat. We seek to determine the uncertainty of the 53 mph figure considering that the manufacturer that calibrated the sensor does not have access to the specific bat it will be used on and must make some assumptions discussed below.

The sensor aims to relate the angular velocity of the knob  $\dot{\omega}_k$  to the angular velocity of the tip  $\dot{\omega}_t$  through

$$\dot{\omega}_t = -\dot{\omega}_k = -\dot{\omega}.$$

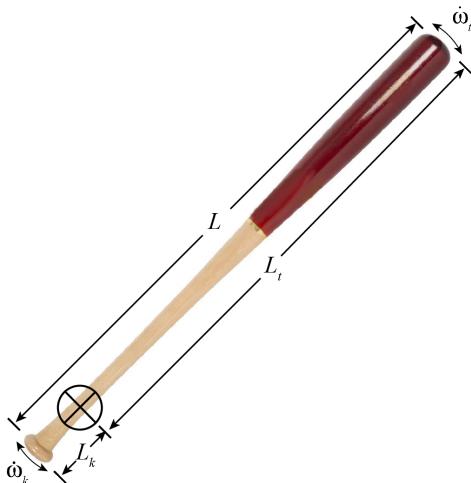


Figure 5.10: Baseball bat showing the relevant lengths and rotation rates.



Figure 5.11: Two 30-inch baseball bats made by different manufacturers standing on their ends.

The tip velocity is this value multiplied by  $L_t$ , or

$$V = \dot{\omega}(L - L_k). \quad (5.18)$$

The bat length  $L$  is specified by the manufacturer to be 30 inches. However, not all 30 inch bats are the same length, as shown in Fig. 5.11. Since the end of the bat may be round or since there may be a plastic cap on the end of the bat that does not contribute to its performance, we assign an uncertainty of  $\pm 0.15$  inches, which is to say  $u_L = 0.15/1.96 = 0.077$  inches.

A major challenge to this measurement is knowing the point where

the player puts their hands,  $L_k$ . The sensor manufacturer assumes the pivot location is 5 inches from the knob, but a sample of 10 players shows a standard deviation in that value of 0.25 inches. This variation may seem random in nature, but note that a single player will use the bat. So,  $u_{L_{k_1}} = 0.25$  inches. The player does not repeat the grip on the bat perfectly each time leading to a random uncertainty of  $r_{L_k} = 0.1$  inches.

The angular velocity sensor has a full-scale output of  $4000^\circ/\text{sec}$ . The manufacturer's specifications state the following:

- Sensitivity Scale Factor  $8.2 \text{ LSB}/(\text{ }^\circ\text{/s})$
- Sensitivity Scale Factor Tolerance  $\pm 2\%$
- Sensitivity Scale Factor Variation Over Temperature for  $-40^\circ\text{C}$  to  $+85^\circ\text{C}$   $\pm 3\%$
- Nonlinearity  $\pm 0.3\%$
- Initial ZRO Tolerance  $\pm 2^\circ/\text{sec}$

Although the specification sheet does not explicitly say so, these are assumed to be percentages of FSO based upon our knowledge of this type of sensor. It is helpful to note that 1% FSO is  $40^\circ/\text{sec}$ . The first two items can be confusing as they address how much the calibration of each unit may vary from the nominal value. When a unit is purchased, the actual sensitivity (based upon a factory calibration) is provided. No uncertainty of this sensitivity (from, for instance, the calibration source) is provided. This may be because other issues (e.g. temperature drift or non-linearity) are much more significant.

The third specification is perhaps important, but no information is provided about how much of this variation occurs in the expected range of use ( $10^\circ\text{C}$  to  $30^\circ\text{C}$ ). If we assume that this variation is linear, we would expect  $3\% \times 20/125 = 0.5\%$ . This is larger than all remaining specified uncertainties. Based upon all of this, we take  $u_{\dot{\omega}_1} = \frac{20}{1.96} \text{ }^\circ/\text{sec} = 0.18 \text{ rad/sec}$ .

The specifications also cite a zero offset drift of  $5^\circ/\text{sec}$ , which is a second source uncertainty  $u_{\dot{\omega}_2} = 5/1.96^\circ/\text{sec} = 0.045 \text{ rad/sec}$ . Electrical noise is also specified, but is small compared to the zero offset.

With two lengths in the DRE, it is tempting to consider the correlation between these measurements, but one comes from the manufacturer and

Table 5.3: The components of the total standard uncertainty of each of the three variables.

Uncertainty source 1	Uncertainty source 2	Total standard uncertainty
$u_{L_{k_1}} = 0.25$	$r_{L_k} = 0.10$	$u_{L_k} = 0.269$
$u_L = 0.077$		$u_L = 0.077$
$u_{\dot{\omega}_1} = 0.18$	$u_{\dot{\omega}_2} = 0.045$	$u_{\dot{\omega}} = 0.185$

the other has to do with a batter's random hand placement on the bat. So we assume no correlated terms. The problem has 3 uncorrelated variables and we will use Eq. 5.18 as the DRE. The total uncertainties for each variable are tabulated in Table 5.3.

Given the nominal values of the lengths and tip speed, the nominal knob rotation rate is 37 rad/sec or  $2120^\circ/\text{sec}$ . We note that this is far below the FSO of the sensor, but that the same sensor is used for longer bats and more mature athletes who generate much larger rotation rates. The derivates are:

$$\begin{aligned}\frac{\partial V}{\partial L_k} &= -\dot{\omega} = -37 \text{ [1/sec]} \\ \frac{\partial V}{\partial L} &= \dot{\omega} = 37 \text{ [1/sec]} \\ \frac{\partial V}{\partial \dot{\omega}} &= L - L_k = 25 \text{ [in]}\end{aligned}\tag{5.19}$$

The TSM equation is

$$u_V^2 = \left( u_{L_k} \frac{\partial V}{\partial L_k} \right)^2 + \left( u_L \frac{\partial V}{\partial L} \right)^2 + \left( u_{\dot{\omega}} \frac{\partial V}{\partial \dot{\omega}} \right)^2.\tag{5.20}$$

The first term is based on an assumed hand placement on the bat. It is not measured, but varies due to each batter's preference. Filling in the values,  $u_V = \sqrt{9.95^2 + 2.85^2 + 4.63^2} = 11.3 \text{ inch/sec} = 0.64 \text{ mph}$  with most of the uncertainty stemming from uncertainty in the pivot point location  $L_k$ . Note that this number represents 68% confidence and should be doubled to achieve 95% confidence.

Let us consider the meaning of this result. For such an inexpensive device to be used in an uncontrolled environment, the uncertainty is

remarkably low at 2% of the reading. For an adult, with a longer bat rotated at a larger speed, the relative uncertainty will be even smaller. Most of the uncertainty is due to an assumption (pivot point location) that involves some uncertainty due to the user. If the user adhered to the assumption made by the designer, the uncertainty would be half as much and be dominated by the rotation sensor.

Now, we turn to this question: If I swung the bat multiple times in an identical manner, would the output vary? To answer this question, we must ask ourselves which of the error sources of each variable are resampled from one swing to the next. Only  $r_{L_k}$  and  $u_{\omega_2}$  represent an error that changes from one swing to the next, so repeating the propagation with only these uncertainties results in  $r_V = 0.22$  mph.

#### Example 5.4: Using High-Speed Images to Find the Acceleration of a Ball

A high-speed camera with a global shutter (meaning it shutters the entire image simultaneously, counter to a cell phone camera) is a powerful measurement tool. Laboratory-grade cameras have a very accurate frame rate, which enhances their utility as a measurement tool.

In this example, we investigate whether a high-speed camera is a suitable tool for measuring the acceleration of a pitched baseball. As a preliminary test, a ball will be dropped in front of such a camera, since we have a good idea of what the acceleration of such an object should be,  $9.81 \text{ m/s}^2$ . The experimental setup is shown in Fig. 5.12.

The ball is dropped at a distance  $X_B$  from the camera. The relationship between pixels and distance is found through a calibration procedure. The distance in pixels between holes of known spacing in a calibration target at  $X_C$  is found. This relationship is only accurate if the ball is dropped at  $X_C$ . If one examines the line of sight from the top edge of the camera sensor to the top edge of the calibration target (indicated with a dashed line), and similarly for the bottom edge, it is clear that the calibration map will overestimate the displacement if  $X_B < X_C$  by an amount that depends on the distance from the center of the image. The error incurred by dropping the ball at  $X_B$  rather than  $X_C$  is zero at the center of the image and scales with the distance (in pixels) from the center of the image. If the vertical sensor size (in pixels) is  $N_y$  and the vertical index of the

sensor is  $m$ , the parallax error from measuring at  $X_B$  can be removed by scaling the  $y'$  position in pixels by

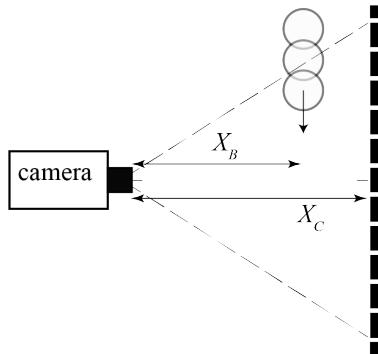


Figure 5.12: Schematic of the experimental setup. A baseball is dropped at a distance  $X_B$  from a high-speed camera. Displacements are converted from pixels to inches by viewing the known spacing of holes in a pegboard at distance  $X_C$ .

$$y' = (m - N_y/2)X_B/X_C. \quad (5.21)$$

This is an example of a known error, and known errors can and should be removed from the result. By doing so, the correction becomes a DRE, since some of the terms are non deterministic, in particular  $m$ ,  $X_B$ , and  $X_C$ . We will now discuss the uncertainty of each of these quantities.

The uncertainty sources are summarized in Table 5.4. Several error sources impact  $X_B$  and  $X_C$ . First, both distances are determined with a tape measure with divisions spaced at 1/16 [in]. While repeated measurements may generate a random result, since this measurement is performed once for each distance, we assign this as an expanded uncertainty of 1/32 [in] for  $X_B$  and  $X_C$ . The standard uncertainty is 1/64 [in]. The sensor location must be estimated due to the presence of the lens, so an uncertainty of 0.5 [in] is assigned to  $X_B$  and  $X_C$ . Also, in the case of  $X_B$ , the ball may not be dropped at exactly the desired location. Preliminary tests show that the distance from the camera to the ball lands has a Gaussian distribution with a standard deviation about 1 [in] due to variations in how it is hand-released. While this variation is random from drop to drop,

Table 5.4: Standard uncertainties of the variables in Eq. 5.23

Source	$m$	$C$	$X_B$	$X_C$	Correlation
Crosscorrelation	3 [pix]	-	-	-	-
Scaling	-	0.5 [pix/in]	-	-	-
Length Resolution	-	-	0.016 [in]	0.016 [in]	-
Cam Sensor Location	-	-	0.5 [in]	0.5 [in]	$0.5 \times 0.5$ [in] <sup>2</sup>
Release Location	-	-	1.0 [in]	-	-
Total	3[pix]	0.5 [pix/in]	1.12 [in]	0.5 [in]	0.25 [in] <sup>2</sup>

for a single test it is a systematic uncertainty. This is much larger than the uncertainty from the tape measure.

Therefore, by the rule of quarters,  $u_{X_C} = 0.5$  [in]. The uncertainty of  $X_B$  has two significant components:

$$u_{X_B} = \sqrt{0.5^2 + 1^2} = 1.12 \text{ [in]}. \quad (5.22)$$

We note that the first of these is correlated to  $X_B$  while the second one is not. The length uncertainty due to the unknown location of the camera sensor is correlated for  $X_B$  and  $X_C$ , so following the guidance of [1],  $u_{X_B X_C} = u_{X_B} u_{X_C}$ .

The value of  $C$  is estimated by finding the number of pixels from one hole in the pegboard to the next. Since the holes are spaced 1 [in] apart,  $u_C = 0.5$  [pix/in].

The vertical index of the ball location, as determined by the cross-correlation algorithm,  $m$ , potentially suffers from myriad uncertainties. Its random uncertainties can be found from the results, as shown below. Uncertainties from cross-correlation are beyond the scope of this text, but based on the width and shape of the correlation peak, we estimate the uncertainty of  $m$  to be 3 [pix].

The  $y'$  can be converted to physical space by dividing by the calibration constant,  $C$ , which is the number of pixels in a inch. The DRE becomes

$$y = \frac{(m - N_y/2)}{C} \frac{X_B}{X_C}. \quad (5.23)$$

Using the Taylor Series Method and Eq. 5.23 as the DRE and noting we are assuming zero uncertainty on  $m$  and the correlation between  $X_B$

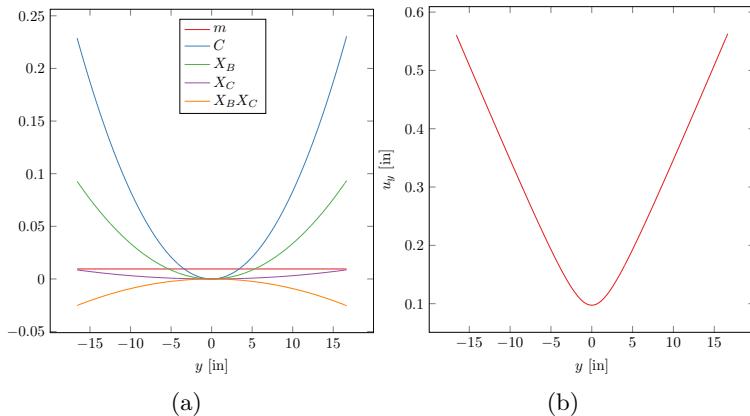


Figure 5.13: (a) The contributions of each of the terms in Eq. 5.24 [in]<sup>2</sup> and (b) the total uncertainty in  $y$  [in].

and  $X_C$ , the uncertainty in  $y$  is

$$u_y^2 = \left( u_m \frac{\partial y}{\partial m} \right)^2 + \left( u_c \frac{\partial y}{\partial c} \right)^2 + \left( u_{X_B} \frac{\partial y}{\partial X_B} \right)^2 + \left( u_{X_C} \frac{\partial y}{\partial X_C} \right)^2 + 2u_{X_B} u_{X_C} \frac{\partial y}{\partial X_B} \frac{\partial y}{\partial X_C} \quad (5.24)$$

We will return to the notation that  $\theta_m = \frac{\partial y}{\partial m}$ , and the derivative values are shown in Table 5.5. While it is often desirable to instead use a relative form of Eq. 5.24 (meaning an equation for  $u_y/y$ ), this would not be a useful equation since  $y$  goes to zero in the center of the field of view. Clearly, Eq. 5.24 is too complicated for any general statements. However, we can plot the contribution of the terms to observe the relative contributions of each and how these vary as the ball moves from the top of the field of view to the bottom, as shown in Fig. 5.13.

Since the correction for perspective is greatest at the edges of the image and minimum at the center, it is not surprising that the uncertainty follows a similar trend. All of the derivatives are a function of  $m$  except the derivative with respect to  $c$ . The most significant source away from the center of the image is the scaling factor  $C$ .

The camera acquires images of the ball at 500 [fps]. An overlay of every

Table 5.5: Derivates of the DRE Eq. 5.23

$\theta_m$	$\frac{X_B}{CX_C}$
$\theta_C$	$-\frac{(m-N_y/2)}{C^2} \frac{X_B}{X_C}$
$\theta_{X_B}$	$\frac{(m-N_y/2)}{CX_C}$
$\theta_{X_C}$	$-\frac{(m-N_y/2)}{C^2} \frac{X_B}{X_C^2}$

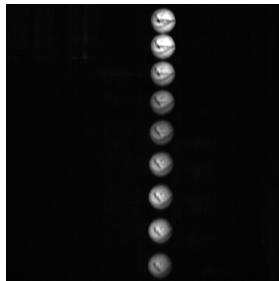


Figure 5.14: Overlay of 9 images of the falling ball. These were acquired at 500 fps and every 10th image is shown.

10th image is shown in Fig. 5.14. Several important features are apparent in this image. First, it is clear that the ball is accelerating downward from the increased spacing between the images. Second, the illumination is not uniform.

The position of the ball as a function of time is determined by cross correlating each image to a white dot that is the same size as the baseball image. Explaining the cross correlation algorithm is beyond the scope of this text, but it returns the most probable spatial shift required in the second image (the white dot) to cause it to appear like the first image (the image of the baseball). The resulting cross-correlation will have a peak near the center of the ball. A script that performs this function was written, and the output is plotted in Fig. 5.15 in which the distances are converted to metric units.

Typical curve fitting functions assume a constant uncertainty, but it is

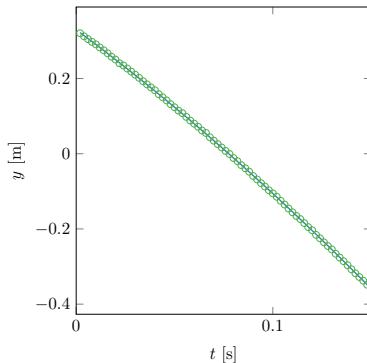


Figure 5.15: The measured location (based on spatial cross correlation) of the ball as a function of time (open symbols) and 2nd order polynomial fit (solid curve).

clear in Fig. 5.13 that the uncertainty in  $y$  varies strongly. This must be taken into account using Eq. 4.15. The polynomial ( $z = C_1 + C_2 t + C_3 t^2$ ) fit to the data that is also included in the plot. The values are found to be  $C_1 = 0.330$  [m],  $C_2 = -3.89$  [m/s], and  $C_3 = -4.75$  [m/s<sup>2</sup>].

We can assess the random uncertainty in the position using Eq. 4.18. It is found to be insignificant compared to the uncertainty shown in Fig. 5.13.

Using Eq. 4.16, we can estimate the standard uncertainties of each of the fit parameters. Using these equations, we estimate the standard uncertainties in this case to be  $u_{C_1} = 7.9 \times 10^{-6}$  [m],  $u_{C_2} = 0.0049$  [m/s], and  $u_{C_3} = 0.19$  [m/s]<sup>2</sup>.

We can find the acceleration by taking two derivatives of the polynomial,

$$a = \frac{d^2 y}{dt^2} = 2C_3 = 9.5 \text{ [m}^2/\text{s}]. \quad (5.25)$$

which is an error of 3% from the known value. We note that this assumes no drag, and that the drag at the final velocity is about 2% of the weight of the ball.

While not directly relevant to the problem at hand, it does provide an opportunity to explore the impact of the number of data points on the uncertainty in a curve fit. We will explore two aspects of the sampling: the sampling rate and the period over which samples are acquired. From Fig. 5.16 it is quite clear that the higher order terms have significantly

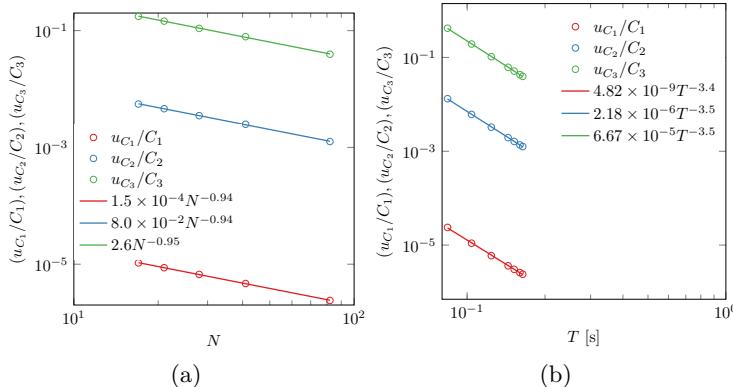


Figure 5.16: The relative uncertainty of the curve fit parameters as a function of (a) The number of images acquired over a fixed period and (b) the length of the period over which image were acquired at a fixed rate. Both plots include a least-squared power law fit.

higher uncertainties ( $u_{C_3}/C_3$  is 4 orders of magnitude larger than  $u_{C_1}/C_1$ ). Fig. 5.16a shows how the number of samples over a fixed time affects the uncertainties while Fig. 5.16b shows how the sampling period affects the uncertainties for a fixed sampling rate. While more data points are better in general, it is clear from the results that the period is more important than the number of samples.

### 5.3 Experiment Planning

The examples discussed thus far are all for cases where the experiment has been completed. Another important use of the propagation of uncertainty is planning an experiment to ensure that the desired level of uncertainty is achievable.

Prior to building any experiment, it is critical to assess whether the desired uncertainty is possible given the budget of the experiment (i.e. the cost of the sensors and the number of data points) and the data reduction equation. The following is to provide answers to questions such as “What is my uncertainty if all quantities can be measured within 5%?”

Coleman and Steele [2] claim that there is little utility to separating

the effects of random and systematic uncertainties for this analysis. One exception would be a case where decreasing the random uncertainty leads to an increase in a systematic uncertainty [27].

For many DREs, it is clear that no one measurement can have an outsized effect on the resultant uncertainty. This is true whenever the derivatives are of similar size. For instance, consider the very common DRE that takes the form

$$y = k(X_1^a)(X_2^b)(X_3^c) \quad (5.26)$$

where  $a$ ,  $b$ , and  $c$  are constants and the exponents can be positive or negative. While 3 variables are shown here, the discussion that follows applies to any number of variables.

Assuming no correlation between the variables, the relative uncertainty using this DRE can be shown to be

$$\frac{u_y}{y} = \sqrt{\left(a \frac{u_{X_1}}{X_1}\right)^2 + \left(b \frac{u_{X_2}}{X_2}\right)^2 + \left(c \frac{u_{X_3}}{X_3}\right)^2}. \quad (5.27)$$

In other words, the relative uncertainty of  $y$  is the root sum square of the relative uncertainties of the measured variables weighted by the exponents. Note that the signs of the exponents are not important due to the squares. In this case, it is a very simple matter to assess the uncertainty in  $y$  given knowledge of the uncertainties of the other variables, and the result are not likely to be surprising.

As an example, consider the drag coefficient:

$$C_d = \frac{2F_d}{\rho V^2 A} \quad (5.28)$$

where  $F_d$  is the drag force,  $V$  is the relative fluid velocity, and  $A$  is the frontal area. This DRE is in the form of Eq. 5.26, and so the relative uncertainty of the drag coefficient may be written as

$$\frac{uC_d}{C_d} = \sqrt{\left(\frac{u_{F_d}}{F_d}\right)^2 + \left(2 \frac{u_V}{V}\right)^2 + \left(\frac{u_A}{A}\right)^2} \quad (5.29)$$

The uncertainty of the drag coefficient will be most affected by the uncertainty of velocity. If all three variables have an uncertainty of 5%, the

uncertainty in the drag coefficient is 12%. If the uncertainty of  $F_d$  and  $A$  are reduced by a factor of 5 to 1%, the uncertainty of the drag coefficient is only reduced to 10%. One should focus on the uncertainty in velocity to reduce the uncertainty in the drag coefficient. For DREs in the form of Eq. 5.26, the focus should be on the variables raised to the largest power.

Some DREs can cause surprisingly high total uncertainties due to a derivative that can go to zero in important parts of the planned experimental space. A DRE feature to be cautious of is one containing a trigonometric function.

For example, alternating current electrical power is the cycle integral of the voltage times the current,

$$P = \int_0^T EI dt, \quad (5.30)$$

where  $T$  is the cycle period, the voltage waveform is  $E = E_1 \sin(2\pi t/T)$ , the current waveform is  $I = I_1 \sin(2\pi t/T + \phi)$ , and  $\phi$  is the phase between voltage and current. By evaluating the integral,

$$P = E_1 I_1 \cos(\phi). \quad (5.31)$$

Applying Eq. 5.10 to this DRE,

$$\frac{u_P}{P} = \sqrt{\left(\frac{u_{E_1}}{E_1}\right)^2 + \left(\frac{u_{I_1}}{I_1}\right)^2 + \tan^2(\phi) u_\phi^2}. \quad (5.32)$$

This tells us that making a measurement at small power levels (when  $\phi$  is near  $90^\circ$ ) becomes problematic. Any finite uncertainty in  $\phi$  will result in unbounded uncertainty in  $P$  in this circumstance.

In some, perhaps rare cases, there may be more than one set of measurements that can be related to the desired quantity. A simple example is DC electrical power, which can be expressed as  $I^2 R$ , where  $I$  is current and  $R$  is resistance, or as  $EI$ , where  $E$  is voltage. Depending on the uncertainties of  $E$ ,  $I$  and  $R$ , one of these DREs may be more advantageous than the other. Examination of Eq. 5.26 reveals that using the first DRE requires a more accurate measurement of current than the second to achieve the same uncertainty in power.

In many texts, planning examples ignore correlated uncertainties. However, in many practical experiments, correlated uncertainties are important

and should be considered at the planning stage. This is especially true when the same sensor and/or data acquisition system is used for multiple measurements. If the derivatives of such correlated measurements are of opposite signs, this will mostly eliminate the effect of their uncertainties. When the derivatives are of the same sign, their uncertainties have increased effects.

In some cases, one has a choice as to the level of correlation between measurements (e.g. separate sensors could be used if less correlation is desirable), meaning that consideration of correlated errors should be part of experimental planning. One should note ratios or differences between quantities using the same sensor to find cases where correlation improves the result. Common examples include temperature differences and before/after mass measurements.

## Summary

If an experimental result is found through a DRE using multiple measurements:

- The correct DRE must be known.
- The uncertainties of each variable must be expressed at the same confidence level.
- The uncertainty of the result is a function of the uncertainties of each variable multiplied by the sensitivity of that variable (its derivative with respect to the result).
- The overall uncertainty may be determined using the Taylor Series Method.
- To do so, one must compute the derivatives of the DRE with respect to each variable or numerically evaluate the derivatives.
- The TSM requires symmetric input distributions and can only produce a symmetric output distribution.
- The MCM is an alternative method that has fewer restrictions but does not provide a closed-form uncertainty solution.

- The MCM has no restrictions on the shape of the input distribution and can produce an asymmetric output distribution (leading to a unique upper and lower uncertainty band).

## Homework

- 1) You are to analyze the uncertainty of  $A$  where  $A = \frac{BC^4}{\pi D}$ . Under what circumstance would TSM and MCM return similar results?
- 2) Which variable in 1) would you want to have the smallest *relative* uncertainty?
- 3) List advantages of MCM compared to TSM.
- 4) Consider the equation  $y = 1.0 - 0.2x + 0.01x^2 + z^{1/2}$ . Determine the uncertainty in  $y(x = 1, z = 1)$  for 2% uncertainty in  $x$  and a 4% uncertainty in  $z$ . Assume no errors are correlated.
- 5) Thermal camera find temperature by measuring the radiative flux from a surface and assuming a known value of the emissivity  $\epsilon$  of the surface:

$$T_s = \left( \frac{q''}{\epsilon \sigma} \right)^{1/4}$$

where  $q''$  is the radiative flux and  $\sigma$  is the Stefan-Boltzmann constant  $5.67 \times 10^{-8} \text{ W/m}^2\text{K}^4$ . Assuming no uncertainty in  $q''$  or  $\sigma$ , find the uncertainty in surface temperature as a function of the  $u_\epsilon$  for various values of  $\epsilon$  between 0 and 1.

- 6) The coefficient of performance of a gas refrigerator is given by

$$COP = \frac{\dot{Q}_L}{\dot{W}_{in}} = \frac{1}{\dot{Q}_H/\dot{Q}_L - 1}.$$

The input power could be measured using a brake that produces a force  $F$

on an arm that is  $L$  long as  $\dot{W}_{in} = FL2\pi\Omega$ , where  $\Omega$  is the rotation rate measured with a tachometer. Both heat transfer rates are determined by measuring the mass flow rate of the air in the heat exchangers and the temperature change across the heat exchanger  $\dot{Q} = \dot{m}C_p\Delta T$ . The table below shows the uncertainties and nominal values of all variables. The 95% confidence uncertainties are shown in the table below. To minimize the uncertainty in  $COP$ , should you use the first or second formula for  $COP$ ? Use TSM assuming no correlated errors. Be careful with units!

Variable	Nominal Value	Total Relative Uncertainty
$L$	2 m	2%
$F$	7.5 N	2%
$\Omega$	955 RPM	2%
$\Delta T_L$	70°C	2%
$\Delta T_H$	100°C	2%
$\dot{m}_H$	0.05 kg/s	2%
$\dot{m}_L$	0.05 kg/s	2%
$C_p$	1 kJ/kgK	2%

- 7) Consider an experiment to determine the convective heat transfer correlation  $h$  for flow over a heated flat plate. Heat flux from a flat plate is

$$q'' = h(T_s - T_\infty) + \epsilon\sigma(T_s^4 - T_{\text{surr}}^4),$$

So, the DRE is

$$h = \frac{\epsilon\sigma(T_s^4 - T_{\text{surr}}^4) - q''}{(T_\infty - T_s)}.$$

The plate emissivity  $\epsilon$  is a material property which depends heavily on aspects of the surface finish that are difficult to measure and can vary over time. Its influence is minimized by keeping its value small through polishing, but its uncertainty is large.

The heat flux is provided through electrical heaters supplied with DC current. The Stefan-Boltzman constant,  $\sigma = 5.67 \times 10^{-8} \text{ W/m}^2\text{K}^4$ , has no significant uncertainty.

Using the nominal values and uncertainties in the table, perform an MCM analysis to determine the uncertainty of  $h$ . Make a plot of the

output distribution. Find a unique upper and lower uncertainty band. Propagate the random uncertainty only using TSM. Compare this value to your MCM result of the total uncertainty. Comment on the meaning of this result.

	$\epsilon$	$q''$	$T_s$	$T_{\text{surr}}$	$T_\infty$
Nominal	0.6	10,000 [W]	350 C	20 C	20 C
$u$	0.05	10W	1°	1°	1°
$r$	0.1	10W	-	-	-



# Chapter 6

## Epilogue

One of the author's primary motivations for writing this text is to expand current thinking on measurement uncertainty to camera-based measurements. The authors were each involved in early efforts to determine the uncertainty of the fluid velocity measurement technique called Particle Image Velocimetry (PIV) [28], [29]. They quickly learned that traditional thinking on uncertainty is not sufficient to tackle these measurements. Camera-based measurements are unique for several reasons:

- The data reduction equation is embodied in thousands of lines of code. There are often dozens of error sources. These error sources may be, in part, a function of the measured quantity while also dependent on other factors. This eliminates the possibility of Taylor's Series Method analysis.
- The uncertainty varies in space and time. It is neither a constant nor a function of the measured variable. This presents difficulty both for determining the uncertainty and also how to use the information that results from an uncertainty analysis.
- Errors may be correlated in space, but not equally over all space.
- These difficulties make it necessary to use the concept of a *ground truth*, or a concomitant measurement of sufficient resolution in time and space and sufficiently low uncertainty to be considered "true" when compared to the technique in question.

## Ground Truth

For traditional measurements, the uncertainty of an instrument is assessed through calibration. As described in Sec. 4.2.3, a systematic uncertainty of the measurement is set to the uncertainty of the calibration source. That uncertainty may be a weak function of one or two more variables (e.g. temperature).

For camera based measurements, errors vary in time and space. In order to establish uncertainties in this complicated environment, it is necessary to have a concomitant measurement of superior accuracy to serve as ground truth. This idea is difficult for many to accept, since the ground truth measurement is not error free. However, it is easy to show that if the ground truth is sufficiently more accurate than the technique being assessed, the ground truth errors are insignificant. It is argued in [29] that if the errors in the ground truth system are smaller than 1/4 of the system under study, they may be considered insignificant. This stems from a similar argument used by Coleman and Steele [2] that when values are to be root summed, any one value that is 1/4 the size of any other is insignificant.

With the “truth” in hand, it is possible to compute errors. The standard deviation of these errors can be used to assess random uncertainties.

# Author's Biographies

**Barton L. Smith** has been a professor in the Mechanical and Aerospace Engineering department at Utah State University in Logan, UT since 2002. While he has always been interested in measurements, his interest in measurement uncertainty began on the day that he computed a negative minor loss factor while working at Los Alamos. His interest intensified when he began performing measurements for the purpose of validating CFD models. Uncertainty of the boundary conditions, inflow and the system response are essential to making such measurements useful.

**Douglas R. Neal** is a senior research engineer at LaVision, Inc where he actively engages in the development of measurement techniques for the scientific community. He regularly interacts with researchers who need to carry out an uncertainty analysis yet they are unsure or unclear on how to proceed. He first became interested in measurement uncertainty during his graduate studies when his PhD committee asked about the uncertainty of data from particle image velocimetry (PIV) data and he realized there that this was not a straightforward task and very little work on that topic had been published in the open literature. He has since become focused on making measurement uncertainty more accessible and understandable in the scientific community.



# Bibliography

- [1] H. W. Coleman and W. G. Steele. *Experimentation, Validation, and Uncertainty Analysis for Engineers*. John Wiley and Sons, Hoboken, NJ, 3rd edition, 2009.
- [2] H. W. Coleman and W. G. Steele. *Experimentation, Validation, and Uncertainty Analysis for Engineers*. John Wiley and Sons, 4th edition, 2018.
- [3] William L Oberkampf and Barton L Smith. Assessment criteria for computational fluid dynamics model validation experiments. *Journal of Verification, Validation and Uncertainty Quantification*, 2(3), 2017.
- [4] B. L. Smith, J. J. Stepan, and D. M. McEligot. Velocity and pressure measurements along a row of confined cylinders. *J. Fluids Eng.*, 129(10):1314–1327, OCT 2007.
- [5] International Organization for Standardization (ISO). *Guide to the Expression of Uncertainty in Measurement*. 1993, ISO, Geneva, Switzerland, Corrected and reprinted, 1995.
- [6] Joint Committee for Guides in Metrology (JCGM). Evaluation of measurement data — guide to the expression of uncertainty in measurement. First Edition, JCGM 100:2008, France, 2008.
- [7] Joint Committee for Guides in Metrology (JCGM). Evaluation of measurement data — an introduction to the “Guide to the Expression of Uncertainty in Measurement”. First Edition, JCGM 104:2009, France, 2009.

- [8] J Chen and J Katz. Elimination of peak-locking error in PIV analysis using the correlation mapping method. *Measurement science and technology*, 16(8):1605, 2005.
- [9] RJ Moffat. Contributions to the theory of single-sample uncertainty analysis. *Journal of Fluids Engineering*, 104(2):250–258, 1982.
- [10] Kun Il Park and Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer, 2018.
- [11] Barton L Smith, Douglas R Neal, Mark A Feero, and Geordie Richards. Assessing the limitations of effective number of samples for finding the uncertainty of the mean of correlated data. *Measurement Science and Technology*, 29(12):125304, 2018.
- [12] Barton L Smith, John Garrett, Patrick Dufour, and Erica Francis. How seams alter boundary layer separation points on baseballs. *Experiments in Fluids*, 65(3):27, 2024.
- [13] Braden J. Limb, Dalon G. Work, Joshua Hodson, and Barton L. Smith. The Inefficacy of Chauvenet’s Criterion for Elimination of Data Points. *Journal of Fluids Engineering*, 139(5), 03 2017. 054501.
- [14] Richard S Figliola and Donald E Beasley. *Theory and design for mechanical measurements*. John Wiley & Sons, 2014.
- [15] Julius S Bendat and Allan G Piersol. *Random data: analysis and measurement procedures*, volume 729. John Wiley & Sons, 2011.
- [16] Andrea Sciacchitano and Bernhard Wieneke. PIV uncertainty propagation. *Measurement Science and Technology*, 27(8):084006, 2016.
- [17] H. Tennekes and J. L. Lumley. *A first course in turbulence*. The MIT Press, 1992.
- [18] Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.
- [19] Dimitris N Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

- [20] Cameron V King and Barton L Smith. Oscillating flow in a 2-d diffuser. *Experiments in fluids*, 51(6):1577–1590, 2011.
- [21] RW Powell, Cho Yen Ho, and Peter Edward Liley. *Thermal conductivity of selected materials*, volume 8. US Department of Commerce, National Bureau of Standards Washington, DC, 1966.
- [22] Marcia L Huber, Richard A Perkins, Daniel G Friend, Jan V Sengers, Marc J Assael, Ifigenia N Metaxa, Kiyoshi Miyagawa, Robert Hellmann, and Eckhard Vogel. New international formulation for the thermal conductivity of H<sub>2</sub>O. *Journal of Physical and Chemical Reference Data*, 41(3):033102, 2012.
- [23] Philip R Bevington and D Keith Robinson. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, 2003.
- [24] Jack Philip Holman. *Experimental methods for engineers*. McGraw Hill, New York, NY, 8th edition, 2012.
- [25] Markus Raffel, Christian E Willert, Fulvio Scarano, Christian J Kähler, Steve T Wereley, and Jürgen Kompenhans. *Particle image velocimetry: a practical guide*. Springer, 2018.
- [26] F.P. Incropera. *Fundamentals of heat and mass transfer*. Number v. 1 in Fundamentals of Heat and Mass Transfer. John Wiley, 2007.
- [27] Rick Cressall, Robert Schaap, Douglas R Neal, Alexander Mychkovsky, John Charonko, and Barton L Smith. Accuracy of volumetric flow rate inflow/outflow measurement by integrating piv velocity fields. *Measurement Science and Technology*, 2020.
- [28] Douglas R Neal, Andrea Sciacchitano, Barton L Smith, and Fulvio Scarano. Collaborative framework for PIV uncertainty quantification: the experimental database. *Meas Sci Technol*, 26, 2015.
- [29] Andrea Sciacchitano, Douglas R Neal, Barton L Smith, Scott O Warner, Pavlos P Vlachos, Bernhard Wieneke, and Fulvio Scarano. Collaborative framework for PIV uncertainty quantification: comparative assessment of methods. *Measurement Science and Technology*, 26(7):074004, 2015.



# Appendices



# **Appendix A**

## **Statistics Tables**

**Table A.1** Tabulation of Two-Tailed Gaussian Probabilities

$\tau$	Prob ( $\tau$ )						
0.00	0.0000	1.00	0.6827	2.00	0.9545	3.00	0.9973002
0.02	0.0160	1.02	0.6923	2.02	0.9566	3.05	0.9977115
0.04	0.0319	1.04	0.7017	2.04	0.9586	3.10	0.9980647
0.06	0.0478	1.06	0.7109	2.06	0.9606	3.15	0.9983672
0.08	0.0638	1.08	0.7199	2.08	0.9625	3.20	0.9986257
0.10	0.0797	1.10	0.7287	2.10	0.9643	3.25	0.9988459
0.12	0.0955	1.12	0.7373	2.12	0.9660	3.30	0.9990331
0.14	0.1113	1.14	0.7457	2.14	0.9676	3.35	0.9991918
0.16	0.1271	1.16	0.7540	2.16	0.9692	3.40	0.9993261
0.18	0.1428	1.18	0.7620	2.18	0.9707	3.45	0.9994394
0.20	0.1585	1.20	0.7699	2.20	0.9722	3.50	0.9995347
0.22	0.1741	1.22	0.7775	2.22	0.9736	3.55	0.9996147
0.24	0.1897	1.24	0.7850	2.24	0.9749	3.60	0.9996817
0.26	0.2051	1.26	0.7923	2.26	0.9762	3.65	0.9997377
0.28	0.2205	1.28	0.7995	2.28	0.9774	3.70	0.9997843
0.30	0.2358	1.30	0.8064	2.30	0.9786	3.75	0.9998231
0.32	0.2510	1.32	0.8132	2.32	0.9797	3.80	0.9998552
0.34	0.2661	1.34	0.8198	2.34	0.9807	3.85	0.9998818
0.36	0.2812	1.36	0.8262	2.36	0.9817	3.90	0.9999037
0.38	0.2961	1.38	0.8324	2.38	0.9827	3.95	0.9999218
0.40	0.3108	1.40	0.8385	2.40	0.9836	4.00	0.9999366
0.42	0.3255	1.42	0.8444	2.42	0.9845	4.05	0.9999487
0.44	0.3401	1.44	0.8501	2.44	0.9853	4.10	0.9999586
0.46	0.3545	1.46	0.8557	2.46	0.9861	4.15	0.9999667
0.48	0.3688	1.48	0.8611	2.48	0.9869	4.20	0.9999732
0.50	0.3829	1.50	0.8664	2.50	0.9876	4.25	0.9999786
0.52	0.3969	1.52	0.8715	2.52	0.9883	4.30	0.9999829
0.54	0.4108	1.54	0.8764	2.54	0.9889	4.35	0.9999863
0.56	0.4245	1.56	0.8812	2.56	0.9895	4.40	0.9999891
0.58	0.4381	1.58	0.8859	2.58	0.9901	4.45	0.9999911
0.60	0.4515	1.60	0.8904	2.60	0.9907	4.50	0.9999931
0.62	0.4647	1.62	0.8948	2.62	0.9912	4.55	0.9999946
0.64	0.4778	1.64	0.8990	2.64	0.9917	4.60	0.9999957
0.66	0.4907	1.66	0.9031	2.66	0.9922	4.65	0.9999966
0.68	0.5035	1.68	0.9070	2.68	0.9926	4.70	0.9999973
0.70	0.5161	1.70	0.9109	2.70	0.9931	4.75	0.9999979
0.72	0.5285	1.72	0.9146	2.72	0.9935	4.80	0.9999984
0.74	0.5407	1.74	0.9181	2.74	0.9939	4.85	0.9999987
0.76	0.5527	1.76	0.9216	2.76	0.9942	4.90	0.9999990
0.78	0.5646	1.78	0.9249	2.78	0.9946	4.95	0.9999992
0.80	0.5763	1.80	0.9281	2.80	0.9949	5.00	0.9999994
0.82	0.5878	1.82	0.9312	2.82	0.9952		
0.84	0.5991	1.84	0.9342	2.84	0.9955		
0.86	0.6102	1.86	0.9371	2.86	0.9958		
0.88	0.6211	1.88	0.9399	2.88	0.9960		
0.90	0.6319	1.90	0.9426	2.90	0.9963		
0.92	0.6424	1.92	0.9451	2.92	0.9965		
0.94	0.6528	1.94	0.9476	2.94	0.9967		
0.96	0.6629	1.96	0.9500	2.96	0.9969		
0.98	0.6729	1.98	0.9523	2.98	0.9971		

**t Table**

cum. prob.	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	<b>0.50</b>	<b>0.25</b>	<b>0.20</b>	<b>0.15</b>	<b>0.10</b>	<b>0.05</b>	<b>0.025</b>	<b>0.01</b>	<b>0.005</b>	<b>0.001</b>	<b>0.0005</b>
two-tails	<b>1.00</b>	<b>0.50</b>	<b>0.40</b>	<b>0.30</b>	<b>0.20</b>	<b>0.10</b>	<b>0.05</b>	<b>0.02</b>	<b>0.01</b>	<b>0.002</b>	<b>0.001</b>
<b>df</b>											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.068	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
<b>Z</b>	<b>0.000</b>	<b>0.674</b>	<b>0.842</b>	<b>1.036</b>	<b>1.282</b>	<b>1.645</b>	<b>1.960</b>	<b>2.326</b>	<b>2.576</b>	<b>3.090</b>	<b>3.291</b>
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	<b>Confidence Level</b>										