Cambridge University Press
978-1-107-02193-8 — Measurement Uncertainty and Probability
Robin Willink
Excerpt
More Information

1

Foundational ideas in measurement

Our analysis gets under way with the identification of concepts in measurement that involve ideas of chance or probability. We begin this important chapter by observing that there are at least two possible views of measurement but that, for the purposes of accommodating notions of probability, one of these views is to be favoured. There follows a discussion about the existence of a 'true value', which seems to be assumed by those concerned with statistical inference but not by measurement scientists. Then we introduce the ideas of error and uncertainty, and emphasize the importance of being clear about the identity of the quantity that is being measured. Subsequently, we invoke the idea that this quantity may be described by a known function of measurable quantities, which is an idea that is foundational both for our analysis and for analysis in international guidelines. There follows a discussion of the central question: *'What does it mean to associate a specified level of assurance, say 95%, with an interval of measurement uncertainty?'* The chapter finishes with a statement of the basic objective of a measurement procedure.

1.1 What is measurement?

A physical quantity subjected to measurement can be called 'the measured quantity'. Yet the result obtained will often be referred to as 'the measured value'. To 'measure a quantity' seems to imply that we are measuring something that exists before the measurement, but to call the figure obtained 'the measured value' implies that we have measured something that did not previously exist. So is the thing measured a quantity that already exists or is it a value that we bring into being? That is, is measurement the *estimation* of the unknown value of the quantity or is it the *declaration* of the known value of the quantity? These questions might seem silly or unnecessary. If so, the reader might have a fixed idea of 'measurement' and might be assuming that everyone shares that idea. But my experience of interacting with other scientists tells me that such confidence would be misplaced: indeed debates about the nature of measurement have a considerable history (e.g. Stevens (1946)).

The noun 'measurement' is not alone in appearing ambiguous in this way. The term 'judgement' seems to suffer similarly – and it might reasonably be said that measurement involves judging the value of a quantity. When witnesses at a trial are asked to judge the speed at which a car was being driven, they are being asked to estimate the speed of the car. But in pronouncing judgement on a party found guilty, the judge will create or prescribe the punishment. So is judgement – like measurement – the estimation of something or the creation of something?

The *International Vocabulary of Metrology* (VIM, 2012, clause 2.1) defines measurement as a

> process of experimentally obtaining one or more **quantity values** that can reasonably be attributed to a **quantity**.

(The words in bold indicate terms defined in that document.) The value attributed to the quantity was not previously associated with that quantity, so this definition is consistent with the idea that measurement is an act of creation or declaration. But it is also consistent with the idea that measurement is an act of estimation. In some situations the value attributed to a quantity can only be understood as an estimate, for example, when measuring a fundamental constant (VIM, 2012, clause 2.11, note 2).

Thus, there seem to be at least two distinct concepts of measurement – one where measurement is the creation or declaration of a known value for the quantity, and the other where measurement is the estimation of an unknown value that is a property of the quantity. Such a fundamental distinction between concepts of measurement will result in differences over the very meaning and role of measurement uncertainty. Therefore this question 'what is measurement?' is addressed at the outset.

Various types of measurement can be defined, but this book is concerned exclusively with the measurement of quantity with a value that exists on a continuous scale, as in the typical measurement problem found in the physical sciences. The view adopted here is that (i) in any well-defined measurement problem of this type we can conceive of a unique unknown value that is the ideal result and that (ii) the act of measurement is the estimation of this unknown value. This view might not be favoured by some readers, but I would ask them to consider the following point, which is argued more fully in the next section: if no such ideal result exists then there is no value to act as a parameter in the statistical analysis. As a result, familiar probabilistic approaches to the analysis of measurement uncertainty become logically invalid.

With this idea of a unique ideal result in mind, I note with gratitude and relief that the *International Vocabulary of Metrology* (VIM, 2012, clause 2.11, note 3) also states that:

> When the **definitional uncertainty** associated with the **measurand** is considered to be negligible compared to the other components of the **measurement uncertainty**, the measurand may be considered to have an "essentially unique" true quantity value. This is the approach taken by the ... [*Guide to the Expression of Uncertainty in Measurement*] and associated documents, where the word "true" is considered to be redundant.

The *measurand* is the 'quantity intended to be measured' (VIM, 2012, clause 2.3). Further support is found in the *Guide to the Expression of Uncertainty in Measurement* (the *Guide*) (BIPM *et al.*, 1995; JCGM, 2008a, clause 3.1.2), which states (with cross-references removed):

> In general, the **result of a measurement** is only an approximation or **estimate** of the value of the measurand ...

The phrase 'approximation or estimate of *the* value of the measurand' suggests that a unique ideal result exists. As we now argue, this concept seems required if the usual ways of evaluating measurement uncertainty are to be meaningful.

1.2 True values and target values

Typically, a discussion of measurement uncertainty involves a notion of 'probability'. Stating a probability requires the existence of a well-defined event or hypothesis. Consider making a probability statement like '$\Pr(X > \theta + 2\sigma) \approx 0.025$', in which θ represents the value of the quantity being measured and X is the random variable for a measurement result. The event inside the parentheses, '$X > \theta + 2\sigma$', must be well defined if this probability statement is to mean anything: the statement makes no sense if θ does not exist, i.e. if θ does not already have a well-defined, albeit unknown, value. Similarly, consider a statement like '$\Pr(\theta > 2.34) = 0.5$', where θ is treated as a random variable. The event '$\theta > 2.34$' must be well defined if this statement is to mean anything – so, once again, θ must have a well-defined, albeit unknown, value.

A statement like '$\Pr(X > \theta + 2\sigma) \approx 0.025$' features in both the classical and Bayesian approaches to the estimation of θ, while a statement like '$\Pr(\theta > 2.34) = 0.5$' appears in a Bayesian analysis. Therefore, for either of those approaches to parameter estimation to be relevant – which are approaches that result in a *confidence interval* and *credible interval* respectively – there must be some unique unknown value θ that can be seen as the ideal result of measurement (see Rabinovich (2007), p. 605). This can be called the *target value* (Murphy, 1961; Eisenhart, 1963). If the existence of such a value is not accepted then no probability statement involving that value can be made and a different approach to the evaluation of measurement data must be sought. (One possible approach based on using probability as a means of outcome prediction rather than parameter estimation is discussed in Section 12.4. The corresponding probability statements involve potential observations only, like the statement 'Pr$(X > 2.34) = 0.5$'.)

Thus, in the measurement of a physical quantity it is helpful to conceive of a unique target value θ. However, this does not mean that we must accept the existence of a single 'true value' for the quantity. Some scientists seem able to point to situations in which, rather than a measurand having a unique unknown value, there is an unknown *interval* of values consistent with the definition of the measurand. If the unknown limits of this interval are θ' and θ'' then we could set $\theta = (\theta' + \theta'')/2$ and, in so doing, define the target value θ for the measurement, as is illustrated in Figure 1.1(a). And some scientists might even consider there to be an unknown *distribution* of values consistent with the definition of the measurand. In such a case we could consider the target value, θ, to be the median of this distribution, as in Figure 1.1(b). Therefore, the absence of a 'true value' does not preclude us from defining a target value.

Perhaps other scientists encounter situations where the value of the quantity being measured fluctuates with time and is representable as $\theta^*(t)$. As a consequence, the idea of a unique target value or true value might seem inappropriate. But what would be the actual purpose of such a measurement? What aspect of the function $\theta^*(t)$ would be the object of study? If we are interested in estimating the whole function $\theta^*(t)$ then we have a unique *target function* – and we are outside

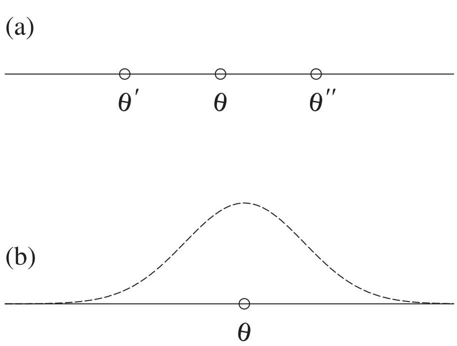

Figure 1.1 Target values θ with two types of 'definitional uncertainty': (a) the target value is the midpoint of an unknown interval; (b) the target value is the median of an unknown distribution.

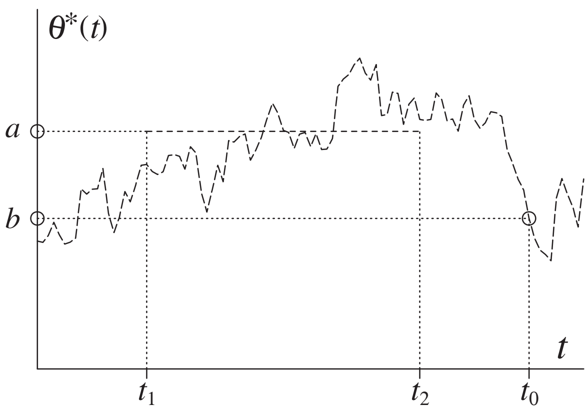

Figure 1.2 Two possible target values when the quantity being measured is fluctuating, $a = (t_2 - t_1)^{-1} \int_{t_1}^{t_2} \theta^*(t)\,dt$ and $b = \theta^*(t_0)$.

the realm of measurement of a scalar, which forms the greater part of this book. However, if we are actually interested in estimating the mean value of $\theta^*(t)$ in some interval $[t_1, t_2]$ then the target value is the scalar $\theta = (t_2 - t_1)^{-1} \int_{t_1}^{t_2} \theta^*(t)\,dt$. And if we are interested in predicting the value of θ^* at some instant t_0 then the target value is the scalar $\theta = \theta^*(t_0)$. These possibilities are illustrated in Figure 1.2.

By such devices, we can define a measurement in a way that the value sought is unique. This enables us to approach the subject of measurement uncertainty using statements of probability that involve this target value: the dispute about the existence or meaning of a 'true value' is put to one side.

1.3 Error and uncertainty

So in any well-defined measurement there is, by some definition, an unknown ideal result of measurement, θ. In that measurement we will obtain a numerical estimate x of θ. If the physical quantity being measured exists on a continuous scale, which is the situation assumed in this book, then θ is unknowable because we cannot measure it with perfect accuracy. Therefore, x cannot be exactly equal to θ, and the quantity $e \equiv x - \theta$ is non-zero, unknown and unknowable. This quantity will feature many times in our analysis, so it requires a name. Although unpopular in some quarters, the usual name for this quantity is the *measurement error*. Thus,

$$\text{measurement error } (e) = \text{ measurement estimate } (x) \; - \; \text{target value } (\theta).$$

The quantity x will also be referred to as the *measurement result*; the terms 'measurement result' and 'measurement estimate' will be used interchangeably.

The measurement error is a signed quantity. Its value is unknown, else we would apply a correction and the error would vanish. However, something will be known about its potential magnitude. This leads us to the idea that a statement of *measurement uncertainty* indicates the potential size of the measurement error. It follows that measurement uncertainty is not the same thing as measurement error: uncertainty means doubt, and doubt remains even if the unknown components of error in the measurement happen to cancel.

The idea of a *standard uncertainty of measurement*, or simply a *standard uncertainty*, is one that has been found helpful. That quantity is often denoted u. Suppose that an appropriate figure of standard uncertainty is 0.03 (in some units). The fact that we would write $u = 0.03$, not $u \approx 0.03$, is indicative of the notion that a standard uncertainty is something determined, not approximated. It is a known estimate of something, not something unknown to be estimated. Thus the error is unknowable but the standard uncertainty is known. One definition of standard uncertainty might be

the best estimate of the root-mean-square value of the error in the measurement procedure.

In Chapter 4, we will find that our understanding of the measurement procedure helps us to view the mean error incurred as being zero. The standard uncertainty of measurement then becomes our best estimate of the standard deviation of the underlying distribution of measurement errors. Figure 1.3 illustrates a situation where the standard uncertainty is equal to this standard deviation.

(Figure 1.3 demonstrates a point of presentation in this book. Where appropriate, unknown quantities will be indicated using hollow markers or dashed lines, while known quantities will be indicated using solid markers and solid lines. In this way,

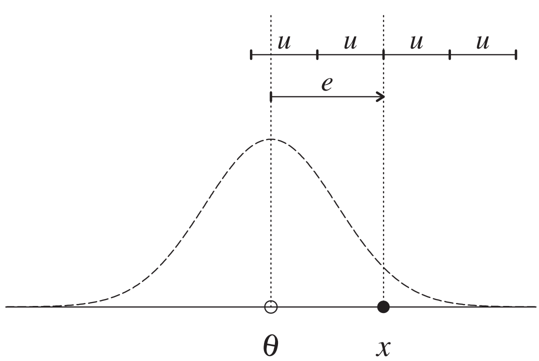

Figure 1.3 Error and standard uncertainty. Dashed line: distribution of potential results when measuring θ; x: known estimate; e: unknown error; u: known standard uncertainty.

we can infer from the body of Figure 1.3 that x is known but that θ and the location of the distribution of potential results are unknown. The same idea is seen at work in Figures 1.1 and 1.2.)

In this book, the concept of a standard uncertainty is not given a great deal of emphasis – and the reader is not obliged to accept the definition of standard uncertainty suggested above. Instead, most of this book is devoted to the task of interpreting and calculating an 'uncertainty interval', which is an output that seems to have greater practical meaning than a standard uncertainty. The concept of an uncertainty interval relates to the idea that, with a specified degree of sureness, we are able to state bounds on the measurement error. Like a standard uncertainty, an uncertainty interval is determined not estimated. This interval depends on the level of sureness we require, and this matter will involve us in a consideration of 'probability'.

The offence of 'error'

For some reason the word 'error' is unpopular in our context. One reason might be distrust of the concept of an ideal result of measurement – without which the terms 'accuracy' and 'error' lack meaning. But, as already indicated, if there is no such thing as an ideal result then the experimentalist has to find an alternative logic when using the concepts of probability.

Another potential reason is that the word 'error' might be thought to imply that a mistake has been made, and the measurement scientist rightly wishes to avoid giving that impression. However, the statistical community finds no offence in the term, and the average citizen does not denigrate market-research companies when they report a 'margin of error' with the result of an opinion poll. Why should it be any different in a measurement problem?

But perhaps the principal cause of the unpopularity of the word 'error' in measurement science is a perception that the techniques of so-called 'error analysis' are inadequate. The techniques of the subject that was known by that name in the middle of the twentieth century do indeed seem limited. The typical text on error analysis would treat all 'random' errors as having arisen from normal distributions, perhaps after invoking the ideas of repeated experimentation and the central limit theorem. More grievously, 'systematic' errors would often be assumed to be either absent or negligible. Thus, the techniques of traditional 'error analysis' are not sufficient for a treatment of the general problem found in measurement science. Yet that is no reason to avoid the concept of error itself. Indeed, the word 'error' seems indispensable: no other word adequately describes the unknown quantity $x - \theta$, and this is a quantity of fundamental importance. Indeed, what is required is a new form of error analysis, a methodology that is designed to overcome the limitations of the older form. Such a methodology is described in this book.

Uncertainty about what?

Uncertainty means doubt, but what are we doubtful about? The measurement estimate x is known: there is no doubt about the identity of x. But the target value θ is unknown, so doubt exists about the value (of) θ. Therefore our uncertainty is about θ, not x, and the language used in relation to measurement uncertainty should reflect this. Simple phrases like 'the uncertainty in x' or 'the uncertainty of x' link the concept of uncertainty to x alone, so such phrases do not seem appropriate. The phrase 'the uncertainty that we associate with x as our estimate of θ' seems correct and informative, but is unwieldy. Instead we will write of the *uncertainty of x for θ*. Acknowledging θ as well as x in the statement of uncertainty is not just a matter of principle: we will see in the next section that the same estimate x can have very different amounts of uncertainty associated with it depending on the definition of θ.

This idea of doubt about θ is to be contrasted with the idea of doubt about the next result when measuring θ, which might be the concept of measurement uncertainty assumed by some readers. Such an understanding would seem influenced by a definition like *the standard uncertainty of measurement is the standard deviation of the set of readings that would be obtained in a very large number of observations* (see Dietrich (1973), p. 8). In this definition, there is no mention of any well-defined quantity being measured, no mention of any target value, and no concept of averaging to improve accuracy. Perhaps most tellingly, there is no mention of the possibility of measurement bias. This might be a definition of measurement uncertainty with some appeal to those who do not accept the concept of a 'true value' or 'target value'. As such, it is a definition with potential relevance to material discussed in Section 12.4.

1.4 Identifying the measurand

The quantity being measured must be well defined: the object being estimated must be clear. This is a fundamental idea, and the reason for not addressing it earlier in the chapter is that the concept of measurement uncertainty plays a key role in our discussion.

Imagine a rod that, apart from the existence of some roughness on one end, has the form of a cuboid. This is illustrated in Figure 1.4. The rod stands on the flat end, and 'the length' of the rod is measured by lowering a probe onto the rough face and recording the vertical coordinate of the point of the probe. This is repeated several

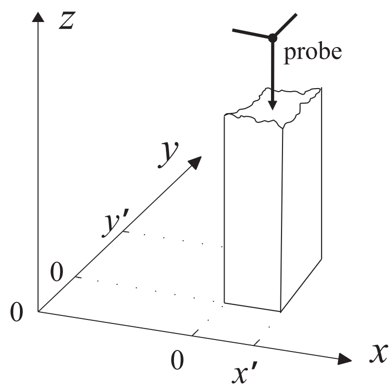

Figure 1.4 Examining lengths of a rod.

times, with the probe being lowered over a different point of the face each time. Let there be n results obtained in total, the jth point on the face being at the coordinates x_j and y_j, and the corresponding vertical distance being $z(x_j, y_j)$. For simplicity, we suppose that this distance is recorded exactly. As might seem reasonable, the arithmetic average of the recorded lengths, $\bar{z} = \sum_{j=1}^{n} z(x_j, y_j)$, is taken as our estimate. But what is the measurement uncertainty to be associated with this figure? The answer depends on the identity of the measurand: what actually is the quantity being estimated?

The figure $\bar{z}$ cannot be called 'our estimate of *the* length of the rod' because the rod has no unique length. It could be said that there are an infinite number of lengths, one for each coordinate (x, y). Yet we have concluded that there is a target value, so something is being estimated and some approximation is being improved by averaging the results of the n measurements. What is being estimated and what approximation is being improved? There are at least two possibilities.

One possibility is that our target is the 'mean length'

$$\mu \equiv \frac{1}{x'y'} \int_0^{y'} \int_0^{x'} z(x, y)\, dx\, dy,$$

which is the volume of the rod divided by the cross-sectional area. This is the mean of the distribution of lengths $z(x, y)$ obtained by associating one length with each infinitesimal element of area $dx\, dy$ on the cross-section. If this distribution of lengths has variance σ^2 and if the coordinates x_j and y_j are chosen randomly then the measurement error $\bar{z} - \mu$ is drawn from a distribution with mean 0 and variance σ^2/n, and the standard uncertainty of measurement will be our best estimate of $\sigma/\sqrt{n}$. Another possibility is that we wish to estimate the length $z_0 \equiv z(x_0, y_0)$ at some unknown point (x_0, y_0) that will subsequently be relevant to our client. Our best estimate of z_0 is again $\bar{z}$, but in this case the error, $\bar{z} - z_0 = (\bar{z} - \mu) + (\mu - z_0)$, is drawn from a distribution with mean 0 and variance $\sigma^2/n + \sigma^2$, so the standard uncertainty of measurement will be our best estimate of $\sigma\sqrt{\{(n+1)/n\}}$.

Therefore – with $\hat{\sigma}$ denoting our estimate of σ – the standard uncertainty of $\bar{z}$ for μ is $\hat{\sigma}/\sqrt{n}$ but the standard uncertainty of $\bar{z}$ for z_0 is $\hat{\sigma}\sqrt{\{(n+1)/n\}}$. The standard uncertainty differs by a factor of $\sqrt{(n+1)}$ depending on whether we wish to estimate the mean length throughout the cross-section or the individual length at an unknown point in the cross-section.

The same idea applies when measuring a fluctuating quantity $\theta^*(t)$, as in Figure 1.2. We might record the value of $\theta^*(t)$ at many instants $t_1, \ldots, t_n$ either to estimate the average level of $\theta^*(t)$ in a well-defined period or to estimate the value of $\theta^*(t)$ at some inaccessible single instant. Suppose the observations are exact, so that the spread in the data reflects only the fluctuation in θ^*. In both cases the measurement estimate will be the sample mean $\sum_{i=1}^{n} \theta^*(t_i)/n$. However, in the first case the standard uncertainty will be $\sqrt{(n+1)}$ times smaller than in the second case.

So the standard uncertainty of measurement is associated not just with the measurement result but also with the identity of the measurand. This is our reason for emphasizing that the target value θ must be acknowledged when quoting a standard uncertainty, as in the phrase 'standard uncertainty of x for θ'.

1.5 The measurand equation

The general measurement problem is more complex than the typical statistical problem of parameter estimation. One reason is that the target value θ is often a known function of many unknown quantities each of which must itself be estimated. This can be expressed as

$$\theta = \mathcal{F}(\theta_1, \ldots, \theta_m), \tag{1.1}$$

where $\mathcal{F}$ is known but each θ_i is unknown. We will call (1.1) the *measurand equation*. Often the quantity physically measured in an experiment is one of the θ_i values, say θ_m. This quantity can be written as

$$\theta_m = \mathcal{G}(\theta, \theta_1, \ldots, \theta_{m-1}), \tag{1.2}$$

and the measurand equation (1.1) is recovered by inversion. We will call (1.2) an *experiment equation*.

Example 1.1 The length of a gauge at 20 °C is measured by relating the length of that gauge to the length of a standard gauge in a comparison conducted at a slightly different temperature. The length of interest l is given by the measurand equation

$$l = \frac{l_s(1 + \alpha_s\lambda) + d}{1 + \alpha\lambda}, \tag{1.3}$$