Chapter 2
Uncertainty

2.1 Why are Measurement Results not Certain?

Suppose that you go to your physician and that he measures your blood pressure. He will probably repeat the measurement a few times and will usually not obtain exactly the same result through such repetitions. This is an example of measurement uncertainty. There are many other examples of this limited *repeatability* which is a good reason for being, to some extent, uncertain about the measurement value. Moreover, even in cases in which the reading is stable, for example when we read the temperature in a room from a wall thermometer, we cannot be totally certain about its value, since a thermometer typically has some "tolerance", for example $\pm 1\,^{\circ}\mathrm{C}$, so that if we read $23\,^{\circ}\mathrm{C}$, we may be quite confident that room temperature will be somewhere between 22 and $24\,^{\circ}\mathrm{C}$, but we would not "bet" on any precise value with a high expectation of winning.

Uncertainty does not only concern simple, everyday measurements: researchers at National Metrology Institutes, such as the National Institute for Standards and Testing (NIST) in USA or the Istituto Nazionale di Ricerca in Metrologia (INRiM) in Italy, spend a considerable amount of time in dealing with uncertainty, even though they work at the highest levels of precision. Measurements in psychophysics are also affected by an even more apparent variability, due to intra- and inter-individual variations.

So it is worth considering uncertainty as an inherent characteristic of measurement. As such, it is has been studied deeply since the beginning of the modern measurement science, and it is therefore useful to give a brief overview of the development of uncertainty theory.

G. B. Rossi, *Measurement and Probability*, Springer Series in Measurement Science and Technology, DOI: 10.1007/978-94-017-8825-0_2,

2.2 Historical Background

2.2.1 Gauss, Laplace and the Early Theory of Errors

The importance of accuracy in measurement has probably been recognised since ancient times. Classical and even scriptural texts warn about the incorrectness of faulty measurement. The concern for reliable measurement units is another side of the problem. Modern scientists have been aware of the need for accurate measurement in order to check scientific theories. Yet the explicit treatment of measurement "errors" was only begun at the beginning of the nineteenth century, when the development of scientific theories and of measuring instruments have required an explicit evaluation of instrumental and measurement performance.

In his *Theoria motus corporum coelestium* [1], Carl Friedrich Gauss (1777–1855) discusses how to obtain estimates of the orbital parameters of heavenly bodies on the basis of a set of observations. He distinguishes between *systematic* and *random* errors. This distinction, already mentioned in his *Theoria motus*, is more clearly expressed in the subsequent *Theoria combinationis observationum erroribus minimis obnoxiae* [2]. Due to the importance of this issue, it is worth reading the original text.

"Certain causes of error—he writes—are such that their effect on any one observation depends on varying circumstances that seem to have no essential connection with the observation itself. Errors arising in this way are called *irregular* or *random*... On the other hand, other sources of error by their nature *have a constant effect on all observations of the same class*. Or if the effect is not absolutely constant, its size varies regularly with circumstances that are essentially connected with the observations. These errors are called *constant* or *regular*". Gauss further observes that "this distinction is to some extents relative and *depends on how broadly we take the notion of observations of the same class*". He explicitly excludes the consideration of systematic (regular, in his terminology) errors in his investigation and warns that "of course, it is up to the observer to ferret out all sources of constant error and remove them". This choice of neglecting systematic errors characterises the classical *theory of errors*, and it is probably its main limitation [3]. We shall see later that the need to overcome this limitation has been the driving force behind the studies on uncertainty in the second half of the twentieth century [4].

To see how Gauss deals with random errors, let us consider the measurement of a single constant quantity x by N-repeated observations. We may model the ith observation by

$$y_i = x + v_i, \tag{2.1}$$

where y_i is the ith observed and recorded value, x is the measurand, which remains constant during the observation process, v_i is the (unknown) value assumed by the

probabilistic (or random) variable v during the ith observation, and $i = 1, 2, ..., N$. The random variable v accounts for the scattering that we observe in the data.[1]

This can also be more compactly expressed in vector notation, as

$$\mathbf{y} = x + \mathbf{v}, \tag{2.2}$$

where $\mathbf{y}$ is a vector of observations, and $\mathbf{v}$ is the vector of random measurement errors.

At this point, Gauss needs an explicit expression for the probability distribution of the errors,[2] p_v, and he thus assumes some properties that correspond to the common understanding of measurement errors. He assumes that p_v is symmetric, maximum in its origin and decreasing on each side of the origin. Furthermore, he assumes that the most probable value for x, once the observations $\mathbf{y}$ have been acquired, is the arithmetic mean of the observed values, since "it has been customary certainly to regard as an axiom the hypothesis that if any quantity has been determined by several direct observations, made under the same circumstances and with equal care, the arithmetic mean of the observed values affords the most probable value, if not rigorously, yet very nearly at least, so that it is always safe to adhere to it". If we denote "the most probable value for x" by $\hat{x}$,[3] this key assumption may be explicated as follows:

$$\hat{x} = \overline{y} \triangleq N^{-1} \sum_i y_i. \tag{2.3}$$

On the basis of this assumption, Gauss could derive the famous distribution named after him. In modern notation, if we introduce the standard normal (Gaussian) distribution, with zero mean and unitary variance defined by

$$\varphi(\xi) = (2\pi)^{-1/2} \exp(-\xi^2/2), \tag{2.4}$$

[1] Henceforth, we need the notion of *probabilistic* or *random* variable (we prefer the former term, although the latter is more common). Though we assume that the reader has a basic knowledge of probability theory, for the sake of convenience, we present a brief review of the probability notions used in this book in Sect. 4.1. Note in particular the notation, since we often use a shorthand one. We do not use any special conventions (such as capital or bold characters) for probabilistic variables. So the same symbol may be used to denote a probabilistic variable or its specific value. For example the probability density function of v can be denoted either as $p_v(\cdot)$ or, in a shorthand notation, as $p(v)$. For notational conventions, see also the Appendix at the end of the book, in particular under the heading "Generic probability and statistics".

[2] A definition of probability distribution, also (more commonly) called the probability density function for continuous variables, is provided in Sect. 4.1.8.

[3] In general the "hat" symbol is used to denote an estimator or an estimated value. If applied to the measurand, it denotes the measurement value.

the result can be compactly expressed as

$$p(v) = \sigma^{-1}\varphi(\sigma^{-1}v), \tag{2.5}$$

where σ is the *standard deviation* [5].

A similar result was reached, in a different way by Pierre-Simon Marquis de Laplace (1749–1827), in his *Théorie analytique des probabilités* [6]. Let us consider the case of repeated measurement once more, and let us still assume that the errors v_i are independent and equally distributed. We now also assume that their distribution $p(v)$ is symmetric about the origin and has a finite support. Let $\hat{x} = \overline{y}$ be the selected estimate for x and

$$e = \hat{x} - x \tag{2.6}$$

the estimation error. Then Laplace showed that e is asymptotically normally distributed with a variance proportional to N^{-1}. In this sense, the normal distribution is regarded as the distribution of the estimation error, for a long series of observations.

It is also possible to consider the problem from another favourable viewpoint, traceable once again to Laplace [7]. Indeed, if we consider the measurement error as deriving from the contribution of a large sum of small independent error sources,

$$v = \sum_j w_j, \tag{2.7}$$

if none of them prevails over the others, the distribution of the resulting error tends to be normal provided that the number of the error sources increases.

In conclusion, the classical measurement error theory, developed mainly thanks to the contributions of Gauss and Laplace, concerns random errors only and results in a probabilistic model, the normal distribution, whose validity can be supported by different arguments.

We will reconsider the measurement error theory at a later stage and will discuss its merits and limitations, and how to overcome them. But we shall now go back to consider the problem of uncertainty from a totally different perspective.

2.2.2 Fechner and Thurstone: The Uncertainty of Observed Relations

The problem of measurement uncertainty was also considered, in around the middle of the nineteenth century, by Fechner in a even more fundamental way [8]. For him, *the only reliable judgements that an observer may express with respect to his sensations are either equality or ordered inequalities* (greater than). His law, in fact, was formulated on the basis of such results. In general, people's responses, which we will call "indications", must be regarded as an expression of *non-deterministic* phenomena, since, for the same pair of stimuli, we may obtain different responses

from different subjects (inter-subjective variability) or even from the same subject, by repeating the test (intra-subjective variability).

A typical experiment in early psychophysics consists in the determination of the just noticeable difference between two stimuli. We already know from Chap. 1 that Fechner's law was developed from such differences. Let us discuss this in further detail. Let ϕ_0 denote the physical intensity of a reference, fixed, stimulus, for example a sound at 1 kHz, with a sound intensity level of 60 dB, and let ϕ be a the variable stimulus of the same kind, having a slightly higher intensity than ϕ_0.[4] Let ψ_0 and ψ be the perceived intensities associated with ϕ_0 and ϕ. Suppose now that we make an experiment with different subjects over repeated trials, in which we wish to determine the minimum value for ϕ that gives rise to a perceivable (positive) variation. In practice, we keep ϕ_0 fixed and, we vary ϕ until the subject listening to both stimuli notices a difference between the two, that is he/she perceives the sensation ψ, associated with ϕ, as being more intense than the sensation ψ_0, associated with ϕ_0.[5] This will not always occur at the same value of ϕ, due to differences in the responses of different people or even to differences in the responses of the same person, when the trial is repeated. The result of one such experiment can therefore be expressed and summarised by the conditional probability[6]

$$P(\psi \succcurlyeq \psi_0 | \phi), \tag{2.8}$$

that is the probability that the sensation ψ is "greater" ($\succcurlyeq$) than the sensation ψ_0. This probability is a function of ϕ, which is varied during the experiment (whilst ϕ_0 is kept fixed), and may qualitatively look as shown in Fig. 2.1.

On the basis of this experimental result, the *differential threshold* can be estimated, conventionally but reasonably, by the value $\delta\phi$ at which $P(\psi \succcurlyeq \psi_0 | \phi_0 + \delta\phi) = 0.75$ [9].

More generally, if we consider two objects a and b and the property ψ associated with them, we can consider the probability $P(\psi_b \succcurlyeq \psi_a)$, or, in a shorthand notation, $P(b \succcurlyeq a)$. The important point here is that *the empirical relation holding between two sensations is recognised as being probabilistic*. This is a somewhat more fundamental perspective than that of the early theory of errors, since *uncertainty is here ascribed to empirical relations rather than to measurement values*. Since empirical relations play a fundamental role in measurement, uncertainty is understood here as affecting the very roots of measurement.

---

[4] We will discuss loudness measurement in some detail in Chap. 8. Readers who are unfamiliar with acoustic quantities may consult the initial section of that chapter for some basic ideas.

[5] In the practical implementation of the experiment, there are different ways of varying the stimulus, either through series of ascending or descending values, or as a random sequence. The variation can be controlled by the person leading the experiment or by the test subject [9, 10] . In any case, such technicalities do not lie within the sphere of this discussion.

[6] For the notion of conditional probability, see Sects. 4.1.1–4.1.3 of Chap. 4, in this book, as well as any good textbook on probability theory [11].

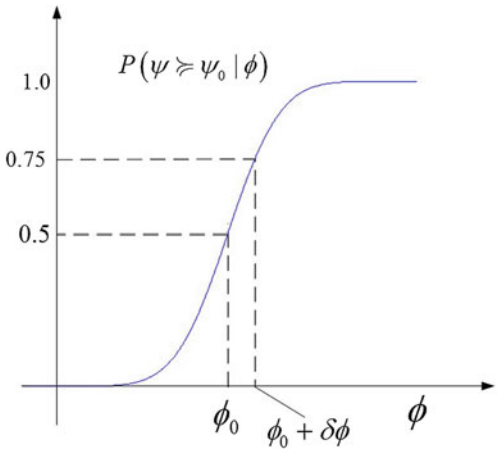

Fig. 2.1 Probability that $\psi \succcurlyeq \psi_0$, as a function of the stimulus ϕ

Once that we have recognised that empirical relations have a probabilistic nature, the challenge is how to represent that in a numerical domain. The solution to this problem will be shown and fully discussed in Chap. 4. For the moment, let us just mention an approach related to the *law of comparative judgement* developed by Luis Leon Thurstone (1887–1955) [12].

Let us then look for a numerical representation of sensations ψ_0 and ψ_1, evoked by objects a and b, respectively, that complies with the empirical evidence that $P(\psi_1 \succcurlyeq \psi_0) = p$, or, equivalently, $P(b \succcurlyeq a) = p$, where p is a probability value, $p \in [0, 1]$.

If we describe ψ_0 and ψ_1 with two independent probabilistic variables, x_a and x_b, whose probability distributions, $p_{x_a}(\psi)$ and $p_{x_a}(\psi)$, are Gaussian, with expected values $\hat{\psi}_0$ and $\hat{\psi}_1$, respectively, and equal variance, σ^2, our condition can be satisfied, provided that

$$\hat{\psi}_1 - \hat{\psi}_0 = z_{10}\sqrt{2}\sigma, \tag{2.9}$$

where z_{10} is such that

$$\int_0^{z_{10}} \varphi(\xi)\mathrm{d}\xi = p - \frac{1}{2}, \tag{2.10}$$

where

$$\varphi(\xi) = (2\pi)^{-1/2} \exp(-\xi^2/2) \tag{2.11}$$

is the standard normal distribution (with zero mean and unitary variance).

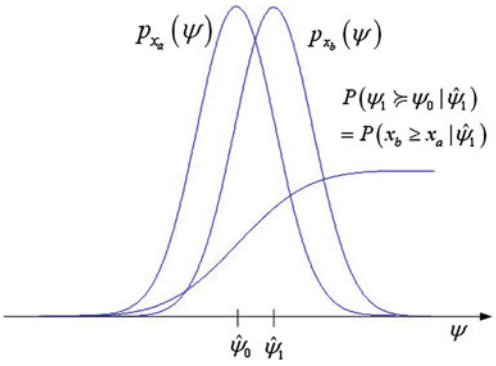

Fig. 2.2 Representation of sensations ψ_0 and ψ_1 by probability distributions on the ψ axis, corresponding to $P(\psi_1 \succcurlyeq \psi_0) = 0.75 = P(x_b \geq x_b)$

Let us briefly show how this result can be obtained. Let us introduce the probabilistic variable $u = x_b - x_a$, which will have mean value $\hat{u} = \hat{\psi}_1 - \hat{\psi}_0$ and standard deviation $\sigma_u = \sqrt{2}\sigma$.[7] Then

$$p = P(x_b \geq x_a) = P(u \geq 0) = \int_0^\infty p(u)\mathrm{d}u = \int_0^\infty (\sigma_u)^{-1} \varphi\left((u - \hat{u})/\sigma_u\right) \mathrm{d}u. \tag{2.12}$$

Making the substitution $v = u - \hat{u}$, we obtain

$$p = \int_{-\infty}^{\hat{u}} (\sigma_u)^{-1} \varphi(v/\sigma_u)\mathrm{d}v = \frac{1}{2} + \int_0^{\frac{\hat{u}}{\sigma_u}} \varphi(\xi)\mathrm{d}\xi. \tag{2.13}$$

Then defining z_{10} as in (2.10), we obtain $z_{10} = \hat{u}/\sigma_u$, from which (2.9) follows.[8] This is illustrated in Fig. 2.2, for $p = 0.75$.[9]

Note that, in this case, $z_{10} = 0.6745$; thus if we denote as $\delta\psi$ the increment in the sensation scale corresponding to a just noticeable increment, $\delta\phi$, in the stimulus, we obtain, approximately,

$$\delta\psi \cong \sigma, \tag{2.14}$$

[7] In fact the variance of the sum (or of the difference) of two independent probabilistic variables equals the sum of their individual variances. Thus, in our case, $\sigma_u^2 = \sigma_{x_b}^2 + \sigma_{x_a}^2 = 2\sigma^2$.

[8] The device of using the abscissae of the standard normal distribution, usually called z-points, is widely used in probability and statistics and, consequently, in psychophysics too.

[9] Interestingly enough, Link notes that Fechner proposed a similar (but not identical) approach, which is very close the signal-detection method of the 1950s. Applying this approach, one would obtain $\hat{\psi}_1 - \hat{\psi}_0 = 2z_{10}\sigma$, instead of the result in (2.9) [13].

which indicates an interesting link between the dispersion of sensorial responses, expressed by σ, and the resolution of the sensation scale, expressed by $\delta\psi$.[10]

As a numerical example, consider again the case of a sound at 1 kHz, with a sound intensity level of 60 dB. The differential threshold value in this case is, roughly, $\delta\phi = 1$ dB [14]. The corresponding loudness values can be obtained by defining the loudness scale for pure tones. We will briefly present such a scale in Chap. 8, Sect. 8.2.2, formulaes 8.17 and 8.11. The measurement unit for that scale is the *sone*. We obtain $\psi_0 = 4$ sone, $\psi_1 \cong 4.3$ sone, that is $\delta\psi = 0.3$ sone, and $\sigma \cong 0.3$ sone.

Thus the two sensations can be represented on a sone scale as probabilistic variables, with mean values $\psi_0 = 4.0$ sone, $\psi_1 \cong 4.3$ sone, and standard deviation $\sigma \cong 0.3$ sone.

2.2.3 Campbell: Errors of Consistency and Errors of Methods

Let us now go back to physical measurement and take a look at Campbell's position. We have already encountered Campbell in the first chapter, as the first (and one of the few) proposers of a comprehensive theory of measurement, at least for physical measurement [15]. In his theory, he also considers measurement errors and distinguishes, as Gauss does, between two kinds of them, which he calls *errors of consistency* and *errors of method*. The former are those that occur when the same measurement is repeated several times under the same conditions and correspond to Gauss's random errors, the latter correspond to systematic errors. It is interesting to see the way he introduces methodical errors: they appear as violations of empirical relations, in particular as violation of "equality", or equivalence, in a more modern language. Equivalence should be transitive, yet in "real" measurement it is often possible to find three objects, a, b and c, such that $a \sim b$ and $b \sim c$, but not $a \sim c$. *How is it possible to reconcile this evidence with the possibility of making fundamental measurements of the quantity for which this happens?*

One way to do this is to consider probabilistic rather than deterministic relations. So the solution to this (fundamental) problem raised but not solved by Campbell comes from an idea that is ultimately traceable to Fechner and Thurstone, that is from the other side of the barricade, in the perspective of the Report of the British Association for the Advancement of Science!

This is further evidence in favour of the need for a unified, interdisciplinary theory of measurement. Prior to discussing probabilistic relations in greater depth, we shall review some subsequent important contributions to the treatment of measurement data.

[10] The resolution of a measurement scale is the minimum variation that can be expressed with that scale (see also the glossary, in the Appendix, at the end of the book).

2.2.4 The Contribution of Orthodox Statistics

Orthodox statistics—the term was coined by E. T. Jaynes (1922–1998) [3]—is a school whose principal exponent was Ronald Aylmer Fisher (1890–1962). During the first part of the twentieth century, he made an important contribution to the development of probabilistic-statistical models by providing a store of methods for their use in conjunction with experimentation [16]. Interestingly enough, this approach makes it possible, in some cases, *to model systematic effects*. To understand how this can be achieved, suppose that we have a measuring instrument that may be affected by a residual (additive) calibration error. If we have just one instrument, the calibration error will give rise to a systematic effect since it will remain constant, at least for some time. But if we have a set of independently calibrated instruments, the calibration error will vary randomly amongst the instruments. Consider now an experiment in which we measure the same fixed quantity x with a set of m independently calibrated measuring instruments of the same type, repeating the measurement n times for each instrument and collecting a total of $N = nm$ observations. The experiment can thus be modelled as follows:

$$y_{ij} = x + \theta_i + v_{ij}, \tag{2.15}$$

where

- $i = 1, \ldots, m$ is the index denoting the instruments,
- $j = 1, \ldots, n$ is the index denoting the repetitions,
- θ_i is a probabilistic variable representing the residual calibration error of each instrument and
- v_{ij} is an array of probabilistic variables, representing random samples from a probabilistic variable, v, that models the random error.

In this framework, the residual calibration error θ gives rise to a systematic error, if we consider the indications of a single instrument as "observations of the same class", whilst it varies randomly if we sample instruments from the class of all the instrument of the same type. From the mathematical point of view, to select a single instrument we fix index i to a constant value, i_0, whilst to sort different instruments, we let it vary within a range from 1 to m. Consider now the following averages: the overall

$$\overline{y} = \frac{1}{N} \sum_{ij} y_{ij}, \tag{2.16}$$

and the average per instrument

$$\overline{y}_i = \frac{1}{n} \sum_{j} y_{ij}. \tag{2.17}$$

Then the measurand x can be estimated as

$$\hat{x} = \overline{y}, \tag{2.18}$$

whilst the systematic deviation of the ith instrument by

$$\hat{\theta}_i = \overline{y}_i - \overline{y}. \tag{2.19}$$

Interestingly enough, in this experiment *it is possible to quantify both the effect of random variations and of systematic deviations.* In fact, the variance of the random variations can be estimated as

$$\hat{\sigma}_v^2 = \frac{1}{N-m} \sum_{ij} (y_{ij} - \overline{y}_i)^2, \tag{2.20}$$

whilst the variance of the calibration error by

$$\hat{\sigma}_\theta^2 = \frac{1}{m-1} \sum_{i} (\overline{y}_i - \overline{y})^2. \tag{2.21}$$

So if it is possible to develop experiments in which the quantities that normally give rise to systematic effects are allowed to vary at random, it is possible to quantitatively evaluate their effect. Unfortunately, this is not the general case in measurement and when this approach is not applicable we have to look for another solution which we will describe at a later stage.

2.2.5 Uncertainty Relations in Quantum Mechanics

A decisive contribution to a deeper understanding of measurement uncertainty came, in the twentieth century, from quantum mechanics [17, 18]. As an example, we briefly mention the basic idea behind the celebrated Heisenberg uncertainty relation. Consider the single-split experiment schematically illustrated in Fig. 2.3.

Suppose we have a beam of electrons impinging on a screen with a thin split. The electrons passing through the split will reach a second, photo-sensitive, screen and form an image on it. If the split is very thin, diffraction will occur, and the image will be wider than the split. We can consider the motion of the electrons passing through the split, as characterised by their position and velocity along the y axis. Interestingly enough, this apparatus will reduce the experimenter's uncertainty as regards position but will increase that concerning velocity. Indeed, before reaching the first screen, the position of the electrons is somewhere in the interval D, whilst their velocity in the y direction is equal to zero. The positions of the electrons passing through the screen lies within the interval d, much smaller than D, but, their velocity v_y is no

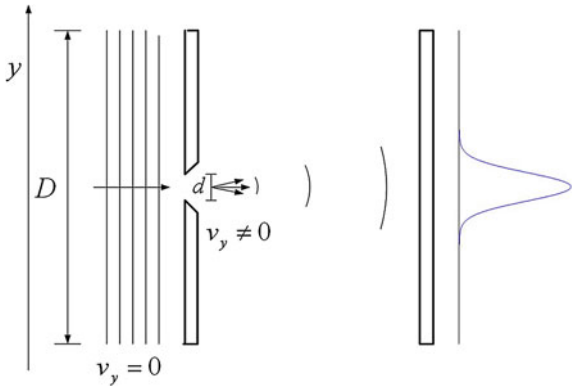

Fig. 2.3 The single-split experiment [18]

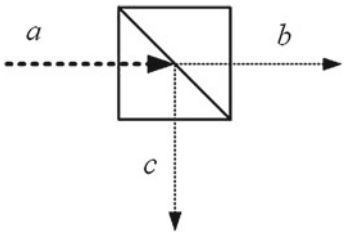

Fig. 2.4 Beam splitter: *a* incident light; *b* transmitted light; *c* reflected light

longer null. If we (informally) denote the uncertainties concerning y and v_y with Δy and Δv_y, respectively, Heisenberg's principle states that

$$\Delta x \Delta v_x \geq \frac{h}{m}, \tag{2.22}$$

where h is Planck's constant and m is the mass of the electron.[11] This is an example of interaction of the measuring system with the object under observation—a system of particles—, which gives rise to a kind of fundamental uncertainty. This suggests that numerical representations in this field must be regarded as inherently probabilistic. In contrast with the classical theory of errors, where, in the absence of systematic effects, measurement precision can be, at least in principle, indefinitely increased, here this is no longer possible, and probabilistic representations are definitely needed.

In fact there is another way, even more important for our purpose, in which quantum mechanics departs from the "classical" perspective of the early theory of errors. To see this in the simplest way, consider a beam splitter, that is an optical device that splits a beam of light in two, as shown in Fig. 2.4.

It is quite easy to describe the macroscopic behaviour of such a device: one half of the incident light is transmitted, whilst the other half is reflected. But now consider a single photon of light: what happens in this case? It cannot be split any further,

[11] This formulation is somewhat qualitative but sufficient for the purpose of this informal discussion.

since it is an elementary, indivisible, entity; it may therefore either pass through or be reflected. This situation can be described in probabilistic terms by assigning a probability of 0.5 to each of these two possibilities. But note now an important difference with respect to the early theory of errors, which was developed within the framework of classical mechanics. In that theory, probability was essentially intended as accounting for our "ignorance" (partial knowledge). Ideally, should we dispose of all the information needed to fully describing the system, we would have no "error". But here, in the new quantum mechanics framework, we cannot describe the behaviour of an elementary item, such as a photon or a particle, better than in probabilistic terms. This makes a big difference. It is often said that in this case probability has a non-epistemic nature ("epistemic" means: related to our state of knowledge).

An important lesson can be learned from quantum mechanics. The approach to uncertainty made in this book will consider measurement as an inherently uncertain process, and we will develop the theory in this perspective. We will also briefly discuss the "nature" of probability in Sect. 4.1.1 and will take a position in that regard.

2.2.6 The Debate on Uncertainty at the End of the Twentieth Century

In the late 1970s, the metrological community recognised the need to reach an internationally agreed way of expressing measurement uncertainty. It also recognised the need to accompany the reporting of measurement results by some quantitative indications of its quality, not only in primary metrology, but also in everyday measurements. In 1978, therefore, the Bureau International des Poids et Mesures (BIPM) carried out an investigation on a large number of laboratories and prepared a recommendation, INC-1 (1980), which was also adopted by the Comité International des Poids et Mesures (CIPM).[12] An international working group was then established, under the guidance of the International Organization for Standardization (ISO), for the purpose of developing a detailed technical Guide. One of the major scientific problems to be faced was the composition of random and systematic effects causing uncertainty. The work of the group was paralleled by intensive scientific debate on these issues. In 1993, the "Guide to the expression of uncertainty in measurement" (GUM) [19, 20] was published. The document had a great impact on both technical and scientific aspects and further stimulated international debate on measurement uncertainty and related topics.

[12] The BIPM and the CIPM are two of the main bodies in the international system of metrology and were established when the Metre Convention was signed (1875). A concise introduction to the organisation of the system is made in Sect. 3.7.4. and additional details on how it works are given in Sect. 10.1.

Let us briefly review some of its main points. Firstly, the GUM recognises, several possible sources of uncertainty, including the following:

1. incomplete definition of the measurand;
2. imperfect realisation of the definition of the measurand;
3. non-representative sampling—the sample measured may not represent the defined measurand;
4. inadequate knowledge of the effects of environmental conditions on the measurement or imperfect measurement of environmental conditions;
5. personal bias in reading analogue instruments;
6. finite instrument resolution or discrimination threshold;
7. inexact values of measurement standards and reference materials;
8. inexact values of constant or other parameters obtained from external sources and used in the data-reduction algorithm;
9. approximations and assumptions incorporated in the measurement method and procedure;
10. variations in repeated observations of the measurand under apparently identical conditions.

Then, addressing uncertainty evaluation, the GUM adopts the paradigm of indirect measurement which has already been mentioned in Chap. 1. In this kind of measurement, the value of the measurand is not obtained directly from the measuring instrument, but by first measuring other quantities that are functionally related to the measurand, and then processing data according to this functional relation. This may be expressed as

$$x = g(\mathbf{z}), \tag{2.23}$$

where x is the measurand, $\mathbf{z}$ a vector of quantities functionally related to the measurand and g a function.[13] We shall call this expression the *(GUM) evaluation model* or *formula*. The quantities appearing in it are treated as probabilistic (or random) variables and their standard deviation, here known as standard uncertainty and denoted with u, is of special interest. Basically the formula allows the uncertainties on the quantities $\mathbf{z}$ to be "propagated" to the measurand x, as we will see in a moment. In turn, these uncertainties may be evaluated on the basis of different pieces of information, which the GUM classifies into two main categories: those coming from a series of observations (type A) and those coming from other sources, such as information provided by the instrument manufacturers, by calibration, by experience, and so on (type B). Note that, in this approach, the focus moves from the type of the uncertainty sources (systematic vs. random) to the type of information concerning them (type A vs. type B). Consequently, it is possible to pragmatically support a common treatment for both of them.

[13] We do not use the GUM's notation here, since we wish to be consistent with the notation used in this book. See the Appendix for further details.

Let us now see how can we apply this approach to the basic case in which we obtain the measurement result directly from a measuring system. We can interpret one of the z_i, for example the first one, as the indication, y, of the measuring system, that is $z_1 = y$, and the remaining z_i as "corrections" that should ideally be applied to correct the effect of the various error sources. The (possible) spread of the indications is accounted for by considering the variability of the probabilistic variable y. The evaluation procedure for the standard uncertainty then proceeds as follows. Since the variables appearing in the evaluation formula are regarded as probabilistic, if $\hat{\mathbf{z}}$ is the expected value[14] of $\mathbf{z}$, that is $\hat{\mathbf{z}} = \mathrm{E}(\mathbf{z})$, $\Sigma_{\mathbf{z}}$ the covariance of $\mathbf{z}$ and $\mathbf{b}$ the vector of the "sensitivities" of x with respect to $\mathbf{z}$, calculated for $\mathbf{z} = \hat{\mathbf{z}}$, that is

$$b_i = \left.\frac{\partial g}{\partial z_i}\right|_{\mathbf{z}=\hat{\mathbf{z}}}, \tag{2.24}$$

then an "estimate" of x may be obtained as

$$\hat{x} = g(\hat{\mathbf{z}}), \tag{2.25}$$

and the standard uncertainty, u, to be associated with $\hat{x}$ is

$$u = \sqrt{\boldsymbol{b}^T \sum_{\mathbf{z}} \boldsymbol{b}}. \tag{2.26}$$

At present, the GUM is an important international reference for the evaluation of measurement uncertainty. Yet, as we have seen, the proposed solution is based on a pragmatic agreement reached within the working group that developed it, rather than on a coherent measurement theory. In this book, we will attempt to do the opposite, that is, derive the rules for evaluating and expressing measurement uncertainty from an overall probabilistic theory of measurement. Due to its importance, uncertainty evaluation will be specifically addressed in Chap. 9. We will also consider the extension of these ideas to all the domains of science, including experimental psychology in particular. Some indications in this sense will be provided in Chap. 8.

2.3 The Proposed Approach

In the above brief historical review, we have learned to distinguish between random variations in observations and systematic effects in the measurement process. We have seen how the former may be modelled according to the classical theory of errors, whilst the latter requires a different approach. Orthodox statistics has provided

[14] In the GUM, the expected value of a quantity is regarded as a "best estimate" of that quantity.

a model for randomising these effects, where practically possible, in order to gain some control on the variables affecting the experiment. Although this is not the general case in measurement, this method is certainly useful, when it is applicable; otherwise a different approach is needed. We have also seen that in psychophysics empirical relations are understood to have a probabilistic character and that in quantum mechanics quantities are regarded as inherently probabilistic. Lastly, we have seen how internationally recognised guidelines are devoted to the evaluation and expression of measurement uncertainty.

In this book, we will develop a general probabilistic approach to measurement that enables uncertainty to be considered and treated in all its forms, in rigorous probabilistic terms.

In Chap. 1, we have seen that in order to measure something a reference scale must first be established and then at least one measuring system based on that scale devised. In dealing with uncertainty, we will follow the same pattern, distinguishing between uncertainty mainly related to the scale and uncertainty mainly related to the measurement process.

2.3.1 Uncertainty Related to the Measurement Scale and to Empirical Relations

A measurement scale is characterised by the empirical relations that can be mapped into corresponding numerical ones. For example, in the case of an ordinal scale, the representation reads

$$a \succcurlyeq b \Leftrightarrow m(a) \geq m(b). \tag{2.27}$$

But this is a deterministic description, since it really implies that, whenever we observe the pair of objects a and b, we always observe either $a \succ b$ or $a \sim b$ or $b \succ a$, and the measurement reflects this state of affairs. Is this what really *always* happens? I suggest readers think of any measurable property of their interest and check if this is the case. I think that making such a statement—a definite relation holds for a, b—is only possible if, intuitively, a and b are "far apart". Instead, if they are "close" to each other it may be, in general, impossible to establish a definite relation between them. To be more precise let us introduce the notion of "comparator", here intended as a device that is capable of establishing an order relation for pairs of elements, with respect to the characteristic under investigation. Let us also consider the notion of "repeatability", this being the ability of a device to produce the same result when operated in the same conditions. Operatively, "same conditions" means that they are undistinguishable for the operator. Repeatability is usually characterised by a standard deviation that quantifies the dispersion of the observations and that can be assessed by a proper calibration test. For example, if we say that a length measuring device has a repeatability of, say, 10 μm, we mean that when repeating

the measurement of an object in undistinguishable conditions, we observe a spread in the instrument indication, with a standard deviation of 10 μm.

So, going back to the issue of comparison, I suggest that if we compare two objects, a and b, whose difference is comparable with the repeatability of the comparator, and we repeat the comparison several times, we may sometimes observe $a \succ b$, sometimes $a \sim b$ and sometimes even $b \succ a$. If this happens, we can say that empirical relations are uncertain and we can describe this situation by means of probability. Very simply, we can assign a probability to each of the possible observed relations, that is

$$P(a \succ b),\ P(a \sim b),\ P(a \prec b), \tag{2.28}$$

satisfying the condition

$$P(a \succ b) + P(a \sim b) + P(a \prec b) = 1. \tag{2.29}$$

We will see later on in this book how to treat the notion, here just presented intuitively, of the probability of a relation in rigorous terms.

To complete this quick look at uncertain relations, we mention that there is another way in which empirical relations may be uncertain. Suppose that we have two equally reliable comparators, C and D, and suppose that, when comparing a and b,

- with C we obtain $a \succ_C b$, whilst
- with C we obtain $a \sim_D b$.

We can interpret this evidence in different ways. We may think that either $a \succ b$ or $a \sim b$ is true and one of the two comparators is wrong, but we do not know which one. Or we may think that the two objects interact with the comparators, in such a way that there are state changes in them, but we are unable to define their states outside these comparisons. Although this uncertainty condition is completely different from the one concerning the issue of repeatability, yet both of them can be described in probabilistic terms. Indeed, in both cases, we can consider $a \succ b$ and $a \sim b$ as uncertain statements characterised by a probability figure.

In Chap. 4, we will see that this yields a probabilistic representation, such as

$$P(a \succcurlyeq b) = P\left(m(a) \geq m(b)\right), \tag{2.30}$$

that replaces formula (2.27). In Chap. 4, we will systematically derive these relations for the scales that are of the greatest interest.

To sum up, I have suggested that the first, in a logical order, sources of uncertainty occurring in measurement may be found in the scale construction phase and that they are related to empirical relations. We may be uncertain about them both due to the lack of perfect repeatability of observations and as a consequence of systematic deviations in what we observe. In both cases, uncertainty can be expressed by probabilistic statements.

2.3.2 Uncertainty Related to the Measurement Process and the Measuring System

The second major part of a theory of measurement concerns the measurement process. We have seen in Chap. 1 that in order to measure something we must first devise a reference scale and then we need a device for comparing unknown objects with the reference scale. We have called such a device a measuring system, and we have also provided a definition for it, as an empirical system capable of interacting with objects incorporating the property under investigation and of producing, as the result of such interaction, signs, on the basis of which it is possible to assign a value to the object to be measured. We can thus model measurement as a process that maps (the properties of) objects in measurement values, that is

$$\hat{x} = \gamma(a), \tag{2.31}$$

where a is an object (considered in respect of a quantity x of its) and $\hat{x}$ is the measurement value that we obtain as the result of the measurement process. Consider now the question of whether such a description is satisfactory or not. Consider what it really implies. It requires that, given any object a, it is always possible to assign to it a measurement value, $\hat{x}$, that exactly describes it.

I do not think that this is generally possible, for reasons similar to those just considered. Again, if we repeat the measurement of the same object several times in equivalent conditions for the experimenter, we may obtain different values,

$$\hat{x}_1, \ldots, \hat{x}_N, \tag{2.32}$$

or, in another scenario, if we measure the same object with two equally reliable devices, R and S, we may repeatedly obtain two different values, one for each system,

$$\hat{x}_R = \gamma_R(a), \quad \hat{x}_S = \gamma_S(a). \tag{2.33}$$

Again it is possible to express such evidence in probabilistic terms. Basically we interpret the *statement* $\hat{x} = \gamma(a)$, where $\hat{x}$ is a number in a set X, as an *uncertain* one, to which a probability can be assigned,

$$P\left(\hat{x} = \gamma(a)\right). \tag{2.34}$$

This probabilistic representation may be interpreted as the probability of obtaining the measurement value $\hat{x}$, for $\hat{x}$ varying in a set of possible values, when object a is measured. This can also be expressed by the conditional probability distribution

$$P\left(\hat{x}|a\right), \tag{2.35}$$

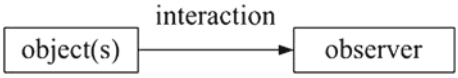

Fig. 2.5 Ideal communication between the object(s) and the observer

where the conditioning event may be modelled as the "random extraction" of object a from the set A. In Chap. 5, we will derive a general expression for such a distribution, based on a general characterisation of the measuring system.

2.3.3 Information Flux Between the Objects(s) and the Observer

Let us now consider a different standpoint that is transversal to the above consideration.

In the case of an observed empirical order relation, $a \succcurlyeq b$, we have a pair of objects, a and b, and a relation between them, established by a comparator. We can thus consider an objects/observer scheme in which the comparator "observes" the empirical relation that holds between the objects. Similarly, in the case of a ternary relation, $a \succcurlyeq b \circ c$, we have three objects, a, b and c, and a device that can establish whether the relation holds or not. So, by extension, we will call *comparator* the device that allows us to establish an empirical relation, and we regard it (including the operator that handles it, in the case of manually operated systems) as an *observer*.

In the case of a measurement process, we have the object to be measured and the measuring system (plus the operator where applicable), in the role of the observer. In both cases, therefore, we can synthetically depict the ideal communication situation as shown in Fig. 2.5.

With respect to this ideal situation, uncertainty sources may affect either the object(s) or the observer or their interaction. The latter hypothesis includes the case, very important in many measurements, in which the measuring system modifies the state of the object.[15] This means that the interaction between object and observer is no longer uni-directional, as in the ideal case,—the object modifies the state of the observer and information is transmitted thanks to this modification—but the observer also modifies the object, and so the state that we actually observe is no longer the original one. All these possibilities are illustrated in Fig. 2.6.

To sum up, I have proposed a taxonomy of uncertainty sources, based on three conceptual coordinates, considering uncertainty either as

- related to empirical relations or to the measurement process,
- referring to random variations or to systematic deviations,

[15] This is usually called "loading effect" in the technical literature.

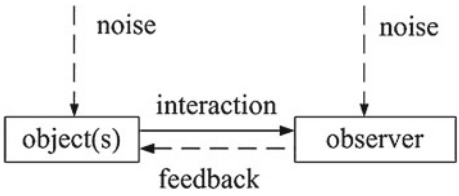

Fig. 2.6 Real communication between the object(s) and the observer, affected by uncertainty sources

- related to the information flux in different ways, that is to say either affecting the object(s) or the observer or their interaction.

I hope that this taxonomy can help in the identification of uncertainty sources as this is the first, and often the most critical, step in uncertainty evaluation. In the second part of the book, we will develop a probabilistic theory, for dealing with uncertainty in general terms, whilst, in the third part, we will discuss some important application issues.

References

1. Gauss, C.F.: Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Hamburg. English edition: (2004) (trans: Davis C.H.). Dover (1809)
2. Gauss, C.F.: Theoria combinationis observationum erroribus minimis obnoxiae. Gottingen. English edition: (1995) (trans: Stewart G.W.). SIAM, Philadelphia (1823)
3. Costantini, D.: I fondamenti storico-filosofici delle discipline statistico probabilistiche. Bollati Boringhieri, Torino (2004)
4. Rossi, G.B.: Probability in metrology. In: Pavese, F., Forbes, A. (eds.) Data Modeling for Metrology and Testing in Measurement Science. Birkhauser-Springer, Boston (2009)
5. Sheynin, O.B.: C. F. Gauss and the theory of errors. Arch. Hist. Exact Sci. **20**, 21–72 (1979)
6. Laplace, P.S.: Theorie analytique des probabilités. Courcier, Paris. In: Oeuvres Complètes de Laplace, vol. 7. Gauthier-Villars, Paris (1812)
7. Sheynin, O.B.: Laplace's theory of errors. Arch. Hist. Exact Sci. **17**, 1–61 (1977)
8. Nowell Jones, F.: History of psychophysics and judgement. In Carterette, E.C., Friedman, M.P. (eds.) Handbook of Perception, vol. 2. Academic Press, New York (1974)
9. Baird, J.C., Noma, E.: Fundamentals of Scaling and Psychophysics. Wiley, New York (1978)
10. Gescheider, G.A.: Psychophysics: The Fundamentals. 3rd edn., Erlbaum, New York (1997)
11. Monti, M., Pierobon, G.: Teoria della probabilità. Zanichelli, Bologna (2000)
12. Thurstone, L.L.: A law of comparative judgements. Psychol. Rev. **34**, 273–286 (1927)
13. Link, S.W.: Rediscovering the past: Gustav Fechner and signal detection theory. Psychol. Sci. **5**, 335–340 (1994)
14. Zwicker, E., Fastl, H.: Psycho-acoustics. Springer, Berlin (1999)
15. Campbell, N.R.: Physics— the Elements. Reprinted as Foundations of Science (1957). Dover, New York (1920)
16. Fisher, R.A.: Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh (1956)
17. Heisenberg, W.: Physics and Philosophy. George Allen and Unwin Edition, London (1959)
18. Ghirardi, G.C.: Un'occhiata alle carte di Dio. Il Saggiatore, Milano (2003)
19. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML: Guide to the Expression of Uncertainty in Measurement. ISO, Geneva, Switzweland. Corrected and reprinted 1995, ISBN 92-67-10188-9 (1993)
20. BIPM: Mutual Recognition. STEDI, Paris (2008)

http://www.springer.com/978-94-017-8824-3

Measurement and Probability
A Probabilistic Theory of Measurement with Applications
Rossi, G.B.
2014, XIII, 286 p. 82 illus., Hardcover
ISBN: 978-94-017-8824-3