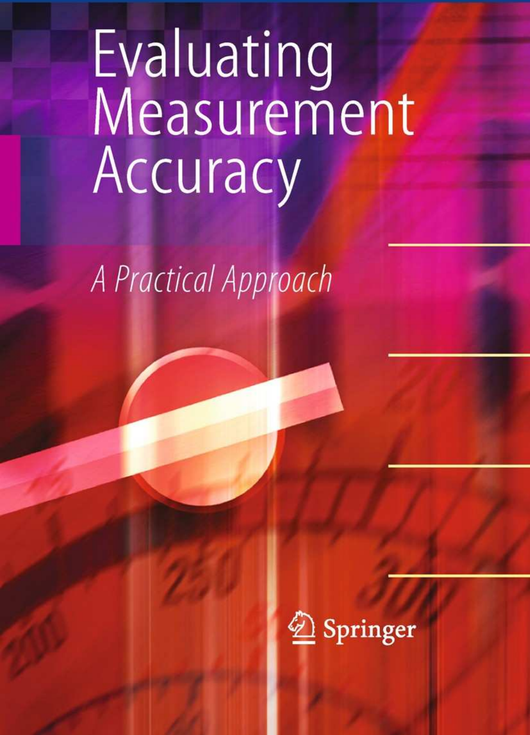
Semyon G. Rabinovich

Evaluating Measurement Accuracy

A Practical Approach



Springer

Evaluating Measurement Accuracy

Semyon G. Rabinovich

Evaluating Measurement Accuracy

A Practical Approach

Springer

Dr. Semyon G. Rabinovich
142 Manchester Drive
Basking Ridge, NJ 07920
USA
semyon_rabinovich@yahoo.com

ISBN 978-1-4419-1455-2 e-ISBN 978-1-4419-1456-9
DOI 10.1007/978-1-4419-1456-9
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009941295

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The goal of this book is to present methods for estimating the accuracy of real measurements, that is, measurements performed in industry, trade, scientific research – wherever the production process, quality control decision, or the interpretation of an experiment depends on measurement accuracy. The necessity for this book arises from the fact that the existing theory of measurement errors contains significant gaps. In particular, the current theory focuses exclusively on multiple measurements and overlooks single measurements. Meanwhile, single measurements are the ones most commonly used in practice. Moreover, the current theory is incomplete even within the scope of multiple measurements. For example, it does not provide answers to such fundamental questions as how to translate the inaccuracy of a measuring instrument into the inaccuracy of a measurement utilizing this instrument, or how to find the full uncertainty of a measurement result, i.e., the uncertainty that reflects both systematic and random errors.

The science of measurements – metrology – entered a period of rapid development several decades ago, prompted by the growth in international trade, globalization of industrial production, demands of medicine and pharmacology, the increased attention to food quality and environment, and other needs of the modern society. However, metrology will not fulfill these needs without removing the blind spots mentioned above. I devoted many years of research filling these gaps. This book generalizes and puts into a coherent whole the results of this effort.

The book develops the general theory of processing experimental measurement data, which addresses the need to obtain the value of a quantity being measured and the accuracy of this estimate. For the first time, this book presents the postulates of the theory of measurements. It introduces the term *measurement inaccuracy* as a general term that reflects measurement uncertainty in some situations and limits of error (or even errors themselves) in others. The book shows the relationship between the accuracy of measuring instruments and measurements utilizing these instruments. It presents methods of estimating the accuracy of both single and multiple measurements. Moreover, it formulates these methods in a systematic and unified way by formulating and utilizing a new perspective that single measurements are the basic type of measurements and multiple measurements represent a series of repeated single measurements. This approach, besides being logical and intuitive, makes accounting for the measuring instruments inaccuracy an inherent part of the calculations of the inaccuracy of the measurement. The book offers well-grounded and practical methods for combining the components of measurement inaccuracy. In particular, it describes how to combine the limits of elementary systematic errors and how to estimate the overall measurement uncertainty accounting for both the systematic and random errors.

As part of the general theory of measurements, the book develops the theory of indirect measurements. For dependent indirect measurements, the book proposes the method of reduction in place of the traditional method based on the Taylor series. This method is more accurate, simpler, and most importantly allows to calculate the confidence limits of the inaccuracy of these measurements, rather than just standard deviation of the measurement result as in the traditional methods. At the same time it removes the need to account for the correlation coefficient, which had been a thorny issue in this area. The book also proposes a new method of transformation for independent indirect measurements. The book further includes a discussion of the applicability of the Bayes' Theorem and Monte Carlo methods in measurement data processing, the topics that have been actively discussed in the metrological research papers.

As a result, this book can serve as a comprehensive reference for data processing of all types of measurements, including single and multiple measurements, dependent and independent indirect measurements, and combined and simultaneous measurements. It includes many concrete examples that illustrate typical problems encountered in measurement practice. Thus, the book encompasses the entire area of measurement data processing, from general theory to practical applications.

The book contains nine chapters. Chapter 1 gives the general introduction to measurements and metrology and outlines major changes that occurred in metrology during the last two decades. Although this chapter is of introductory nature, it presents some important general perspectives on the subject. In particular, it includes a classification of measurements and measurement inaccuracy and formulates postulates of the theory of measurements.

Chapter 2 is devoted to measuring instruments. It describes conventional methods of representing their metrological characteristics as well as the methods of controlling these characteristics through calibration or verification. The chapter also analyzes errors of large numbers of instruments of several types and shows that the distribution functions of these errors are usually nonstationary.

Chapter 3 contains basic statistical methods of experimental data processing. These methods are directly applicable to the idealized multiple measurements. They are also necessary when using statistical models of elementary measurement errors and for obtaining confidence intervals in the course of measurement uncertainty calculations.

Chapter 4 is devoted to direct measurements. It presents a step-by-step procedure for the calculations of the inaccuracy of single measurements. The calculation of uncertainty of a multiple measurement is then derived as a summation of the inaccuracy of the underlying single measurement with the random error of the multiple measurement, which is estimated from the repeated single measurements. The chapter describes a new summation method and the advantages of the new method over the known methods of summating systematic and random errors. Finally, the chapter briefly describes nonparametric and robust methods for processing direct measurement data.

Chapter 5 presents the theory of indirect measurements. In particular, it describes two new methods. The first one is the method of reduction, which handles indirect measurements with dependent arguments. This method, which we proposed previously but which is not yet widely known, is the first to produce reliable estimates of uncertainty of these types of measurements. At the same time, it eliminates the need to calculate the correlation coefficients – a major stumbling block in these measurements. The second method is the method of transformation for indirect measurements with independent arguments, which compliments the known methods. This chapter also applies the new general method for the summation of systematic and random errors from Chap. 4 to indirect measurements, thus removing the need to use the Monte Carlo method with its known limitations (the reliance on unknown distribution functions and the complexity of implementation) for this purpose.

Chapter 6 treats simultaneous and combined measurements, using the well-known least-squares method, which is commonly applied for these measurement types.

Chapter 7 contains methods for combining measurement data or measurement results. This problem arises when the same measurand is measured in different laboratories, and the final result should reflect all these measurements. Along with a traditional solution, which takes into consideration only random errors, Chap. 7 includes a method accounting for the systematic errors as well.

Chapter 8 includes a number of concrete examples of measurement data processing and evaluating measurement accuracy. The book is targeted for practical use, and these examples can serve as specific blueprints for addressing typical measurement data processing needs faced by experimenters.

Finally, Chap. 9 presents concluding remarks, including our thoughts on the new "International Vocabulary of Metrology – Basic and General Concepts and Associated Terms" [1] and analysis of the drawbacks of the "Guide to the Expressions of Uncertainty in Measurement" [2].

This book is intended for anyone who is concerned with measurements in any field of science or technology, who design technological processes and chooses instruments with appropriate accuracy as part of their design, and who design and test new measuring devices. This book should also be useful to university students pursuing science and engineering degrees. Indeed, measurements are of such fundamental importance for modern science and engineering that everyone in these fields must know the basics of the theory of measurements and especially how to evaluate their accuracy.

The book assumes reader's familiarity with mathematical statistics, basic calculus, and, in part of Sect. 2.5, control theory. A reader without control theory background can skip this part of Sect. 2.5 without affecting the understanding of the rest of the book.

In conclusion, I would like to thank Dr. Abram Kagan, Professor at the University of Maryland, College Park, for many years of collaboration and friendship. This book benefited from our discussions on various mathematical problems in metrology. I would also like to thank Dr. Ilya Gertsbach, Professor at the Ben Gurion University of Beersheva (Israel), for our discussions over the theory of independent indirect measurements. I would like to express my special gratitude to my son, Dr. Michael Rabinovich, Professor at Case Western Reserve University. He provided support and assistance throughout my work on this book from editing the proposal for publication to discussing new results and the presentation to editing the whole book. This book would not be possible without his help.

Basking Ridge, NJ Semyon G. Rabinovich

Contents

1 General Concepts in the Theory of Measurements 1
- 1.1 Basic Concepts and Terms 1
- 1.2 The Basic Metrological Problems 4
- 1.3 New Forms of International Cooperation in Metrology 10
- 1.4 Postulates of the Theory of Measurements 12
- 1.5 Classification of Measurements 18
- 1.6 Classification of Measurements Errors 22
- 1.7 General Approach to Evaluation of Measurement Inaccuracy 25
- 1.8 Presentation of Measurement Results 26

2 Measuring Instruments and Their Properties 31
- 2.1 Types of Measuring Instruments 31
- 2.2 Metrological Characteristics of Measuring Instruments 33
- 2.3 Rating of the Errors of Measuring Instruments 36
- 2.4 Dynamic Characteristics of Measuring Instruments 49
- 2.5 Calibration and Verification of Measuring Instruments 53
- 2.6 Designing a Calibration Scheme 58
- 2.7 Statistical Analysis of Measuring Instrument Errors 65

3 Statistical Methods for Experimental Data Processing 71
- 3.1 Methods for Describing Random Quantities 71
- 3.2 Requirements for Statistical Estimates 75
- 3.3 Evaluation of the Parameters of the Normal Distribution 76
- 3.4 Elimination of Outlying Data 80
- 3.5 Construction of Confidence Intervals 82
- 3.6 Testing Hypotheses about the Form of the Distribution Function 87
- 3.7 Testing for Homogeneity of Samples 89
- 3.8 Robust Estimates 95
- 3.9 Application of the Bayes' Theorem 99

4 Direct Measurements 101
- 4.1 Relation Between Single and Multiple Measurements 101
- 4.2 Classification of Elementary Errors 104

4.3 Modeling of Elementary Errors 108
4.3.1 Absolutely Constant Errors 109
4.3.2 Conditionally Constant Errors 109
4.3.3 Purely Random Errors 110
4.3.4 Quasirandom Errors 110
4.4 Composition of Uniform Distributions 111
4.5 Methods for Precise Measurements 117
4.6 Accuracy of Single Measurements Using Measuring Instruments Under Reference Conditions 119
4.7 Accuracy of Single Measurements Using Measuring Instruments Under Rated Conditions 122
4.8 Accuracy of Multiple Measurements 128
4.9 Comparison of Different Methods for Combining Systematic and Random Errors 136

5 Indirect Measurements 141
5.1 Terminology and Classification 141
5.2 Correlation Coefficient and Its Calculation 142
5.3 The Traditional Method of Experimental Data Processing 145
5.4 Merits and Shortcomings of the Traditional Method 149
5.5 The Method of Reduction 152
5.6 The Method of Transformation 154
5.7 Total Uncertainty of Indirect Measurements 161
5.8 Accuracy of Single Indirect Measurements 166
5.9 Accuracy of a Single Measurement with a Chain of Instruments 167
5.10 Application of the Monte Carlo Method 169

6 Combined and Simultaneous Measurements 173
6.1 General Remarks About the Method of Least Squares 173
6.2 Measurements with Linear Equally Accurate Conditional Equations 175
6.3 Measurements with Linear Unequally Accurate Conditional Equations 178
6.4 Linearization of Nonlinear Conditional Equations 180
6.5 Examples of the Application of the Method of Least Squares 181
6.6 General Remarks on Determination of the Parameters in Formulas from Empirical Data 186
6.7 Construction of Transfer Functions of Measuring Transducers 188

7 Combining the Results of Measurements 193
7.1 Introductory Remarks 193
7.2 Theoretical Principles 193
7.3 Effect of the Error of the Weights on the Error of the Weighted Mean 197

7.4 Combining the Results of Measurements with Predominately Random Errors 199
7.5 Combining the Results of Measurements Containing Both Systematic and Random Errors 201
7.6 Combining the Results of Single Measurements 204

8 Examples of Measurements and Measurement Data Processing 209
8.1 Voltage Measurement with a Pointer-Type Voltmeter 209
8.1.1 A Priori Estimation of Measurement Inaccuracy 209
8.1.2 Universal Estimation of Measurement Inaccuracy 212
8.1.3 Individual Estimation of Measurement Inaccuracy 214
8.2 Voltage Measurement with a Potentiometer and a Voltage Divider 215
8.3 Comparison of Mass Measures 220
8.4 Measurement of Electric Power at High Frequency 221
8.5 An Indirect Measurement of the Electrical Resistance of a Resistor 222
8.5.1 Application of the Traditional Method 222
8.5.2 Application of the Method of Reduction 225
8.6 Measurement of the Density of a Solid Body 226
8.6.1 Application of the Traditional Method 226
8.6.2 Application of the Method of Transformation 229
8.7 Measurement of Ionization Current 233
8.8 Measurement of the Activity of a Radioactive Source 235

9 Conclusion 239
9.1 Measurement Data Processing: Past, Present, and Next Steps 239
9.2 Remarks on the International Vocabulary of Metrology 241
9.3 Drawbacks of the "Guide to the Expression of Uncertainty in Measurement" 244

Appendix 249

Glossary 255

References 261

Index 265

About the Author

Semyon Rabinovich was Director of the Laboratory for Theoretical Metrology in St. Petersburg [then Leningrad] from 1965 to 1980, at which time he emigrated from the Soviet Union. He has since consulted on nuclear safety instrumentation and on other matters. Dr. Rabinovich is credited with developing an entirely new branch of metrology: physical metrology, in which theory is based on instrumentation rather than pure mathematics. An earlier book he authored, called *Measurement Errors and Uncertainties* (third edition published by Springer in 2005) is more concerned with theory. This book, though bearing some resemblance to his earlier book, is much more practical than theoretical.

Abbreviations

Names of Organizations and Institutions

APLAC	Asia Pacific Laboratory Accreditation Cooperation
BIPM	International Bureau of Weights and Measures
CGPM	General Conference of Weights and Measures
CIPM	International Committee of Weights and Measures
EURACHEM/CITAC	A network for Cooperation in International Traceability in Chemical Measurements
EUROMET	European Collaboration in Measurement Standards
IEC	International Electrotechnical Commission
ILAC	International Laboratory Accreditation Cooperation
JCGM	Joint Committee for Guides in Metrology
NAFTA	North America Free Trade Agreement
NCSL	National Conference of Standard Laboratories
NCSLI	National Conference of Standard Laboratories International
OILM	International Organization of Legal Metrology

Titles of Documentations and Other Abbreviations

AC	Alternating current
DC	Direct current
EMF	Electromotive force
GUM	Guide to the expression of uncertainty in measurement
SI	International system of units
VIM	International vocabulary of metrology – basic and general concepts and associated terms

Chapter 1
General Concepts in the Theory of Measurements

1.1 Basic Concepts and Terms

The theory of measurement accuracy is a branch of metrology – the science of measurements. In presenting the theory we shall adhere, whenever possible, to the terminology given in the *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* [1]. We shall discuss the terms that are most important for this book.

A measurable quantity (briefly – quantity) is a property of phenomena, bodies, or substances that can be defined qualitatively and expressed quantitatively. The first measurable quantities were probably length, mass, and time, i.e., quantities that people employed in everyday life and these concepts appeared unconsciously. Later, with the development of science, measurable quantities came to be introduced consciously to study the corresponding laws in physics, chemistry, and biology.

The term *quantity* is used in both the general and the particular sense. It is used in the general sense when referring to the general properties of objects, for example, length, mass, temperature, or electric resistance. It is used in the particular sense when referring to the properties of a specific object: the length of a given rod, the electric resistance of a given segment of wire, and so on. The principal feature of quantities in the context of this book is that they can be measured. A *measurand* is a quantity intended to be measured.

Measurement is the process of determining the value of a quantity experimentally with the help of special technical means called *measuring instruments.*

The *value of a quantity* is the product of a number and a unit adopted for these quantities. It is found as the result of a measurement. This definition can be expressed in the form of the equation:

$$Q = q[Q],$$

where Q is the value of the measurand, $[Q]$ is a unit adopted for the kind of quantity represented by the measurand, and q is the number showing how many of these units constitute the magnitude of the measurand. This equation is sometimes called the *basic measurement equation*. Note that the unit is not indicated if the measurand is dimensionless.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_1,

The basic measurement equation reflects the general objective of a measurement: to express with a number a property of an object or natural phenomenon. Thus measurements allow us to use mathematics in our practical activities and in the exploration of nature.

The definitions presented above underscore three features of measurement:

1. The result of a measurement must always be a concrete denominated number expressed in sanctioned units of measurements. The purpose of measurement is essentially to represent a property of an object by a number.
2. A measurement is always performed with the help of some measuring instrument; measurement is impossible without measuring instruments.
3. Measurement is always an experimental procedure.

The *true value of a measurand* is the value of the quantity, which, if known, would ideally reflect, both qualitatively and quantitatively, the corresponding property of the object according to the purpose of the measurement.

Measurement accuracy reflects the closeness between the measurement result and the true value of the measurand. Measuring instruments are created by humans, and every measurement on the whole is an experimental procedure. Therefore, results of measurements cannot be absolutely accurate.

Accuracy is a "positive" characteristic of the measurement, but in reality it is expressed through a dual "negative" characteristic – inaccuracy – of the measurement. The inaccuracy reflects the unavoidable imperfection of a measurement. The inaccuracy of a measurement is expressed as the deviation of the measurement result from the true value of the measurand (this deviation is called the measurement error) or as an interval that covers the true value of the measurand. We will call the half-width of this interval *uncertainty* if it is obtained as a confidence interval (i.e., the interval that covers the true value with a certain probability) and *limits of error* if it has no relation with probabilities. We shall return to these terms many times later in this book.

The true value of a measurand is known only in the case of calibration of measurement instruments. In this case, the true value is the value of the measurement standard used in the calibration, whose inaccuracy must be negligible compared with the inaccuracy of the measurement instrument being calibrated.

A measurement error can be expressed in absolute or relative form. The error expressed in the absolute form is called the absolute measurement error. If A is the true value of the measurable quantity and $\tilde{A}$ is the result of measurement, then the absolute measurement error is $\zeta = \tilde{A} - A$. The absolute error can be identified by the fact that it is expressed in the same units as the measurable quantity. Absolute error is a quantity and its value may be positive or negative. One should not confuse the absolute error with the absolute value of that error. For example, the absolute error -0.3 mm has the absolute value 0.3.

The error expressed in relative form is called the relative measurement error. The relative error is the error expressed as a fraction of the value of the measurand: $\varepsilon = (\tilde{A} - A)/A$. Relative errors are normally given as percent and sometimes per thousand (denoted by ‰). Very small errors, which are encountered in the most precise measurements, are customarily expressed directly as fractions of the measured quantity, given in parts per million (ppm).

In most cases, however, the true value of the measurand is unknown, and the inaccuracy is expressed as an interval covering the true value. As mentioned above, the boundaries of this interval are the uncertainty or limits of error, depending on whether or not the interval was calculated using a probabilistic approach. The interval limits are specified as the offsets from the measurement result; just like measurement errors, these limits can be expressed in the absolute or relative form.

We should note that the above-mentioned equation for the absolute error is often presented as the general definition of measurement error [1, 2, 6, 10].

From our discussion, it should be clear that this definition narrows the meaning of the term measurement error.

The absolute measurement error or uncertainty, depends in general on the value of the measured quantity, and for this reason, it is not a suitable quantitative characteristic of measurement accuracy. Relative errors or uncertainties do not have this drawback. For this reason, measurement accuracy can be characterized quantitatively by the inverse of the relative error or uncertainty expressed as a fraction (not as a percentage) of the measured quantity. For example, if the limits of error of a measurement are $\pm 2 \times 10^{-3}\% = \pm 2 \times 10^{-5}$, then the accuracy of this measurement will be 5×10^4. Note that the accuracy is expressed only as a positive number.

Although it is possible to introduce in this manner the quantitative characteristic of accuracy, in practice, accuracy is normally not estimated quantitatively and it is usually characterized indirectly with the help of the measurement error or the uncertainty of measurement.

The quality of measurements that reflects the closeness of the results of measurements of the same quantity performed under the same conditions is called the *repeatability of measurements*. Good repeatability indicates that the random errors are small.

The quality of measurements that reflects the closeness of the results of measurements of the same quantity performed under different conditions, i.e., in different laboratories (at different locations) and using different equipment, is called the *reproducibility of measurements*. Good reproducibility indicates that both the random and systematic errors are small.

Uniformity of measuring instruments refers to the state of these instruments in which they are all graduated in the established units and their errors and other relevant properties fall within the permissible limits. *Unity of measurements* refers to a common quality of all measurements performed in a region (in a country, in a group of countries, or in the world) such that the results of measurements are expressed in established units and agree with one another within the estimated limits of error or uncertainty.

Uniformity of measuring instruments is a necessary prerequisite for unity of measurements. However, the result of a measurement depends not only on the quality of the measuring instrument employed but also on many other factors, including human factors (if measurement is not automatic). For this reason, unity of measurements in general is the limiting state that must be strived for, but which, as any ideal, is unattainable.

1.2 The Basic Metrological Problems

Comparison is an age-old element of human thought, and the process of making comparisons lies at the heart of measurement: Homogeneous quantities characterizing different objects are identified and then compared; one quantity is taken to be the unit of measurement and all other quantities are compared with it. This is how *measures*, i.e., objects that exhibit quantities of unit size (or the size of a known number of units) came about.

At one time, numerous independent units and measures were used in different regions; even different cities each had their own units and independent measures. Then it became necessary to know how different measures of the same quantity type were related, in order to unify measurements across regions. This problem gave birth to the study of measures, which later turned into the science of measurements – metrology.

But the content of metrology, as that of most sciences, is not immutable. Especially profound changes started in the second half of the nineteenth century, when industry and science developed rapidly and, in particular, electrical technology and instrument building began. Measurements were no longer merely a part of production processes and commerce; they became a powerful means of gaining knowledge – they became a tool of science. The role of measurements has increased dramatically today, in connection with the rapid development of science and technology in the fields of nuclear research, space, electronics, and so on.

The development of science and technology, contacts among peoples, and international trade has prompted many countries to adopt the same units of physical quantities. The most important step in this direction was the signing of the Metric Convention [(Treaty of the Meter), 1875]. This act had enormous significance not only with regard to the dissemination of the metric system, but also with regard to unifying measurements throughout the world by means of the creation of international measurement standards. The Metric Convention and the institutions created by it – the General Conference on Weights and Measures (CGPM), the International Committee of Weights and Measures (CIPM), and the International Bureau of Weights and Measures (BIPM) – continue their important work today. In 1960, the General Conference on Weights and Measures adopted the international system of units (SI) [1, 3]. Most countries now use this system.

The range of topics encompassed by modern metrology is shown in the block diagrams in Fig. 1.1.

While many of the listed topics are self-explanatory, several warrant further examination. We expand on these topics below, beginning with some blocks in the diagram of Fig. 1.1d.

1. The Study of Measurable Quantities and their Units

Measurable quantities are introduced in different fields of knowledge, in physics, chemistry, biology, and so on. The rules for introducing and classifying them and for forming systems of units and for optimizing these systems cannot be addressed

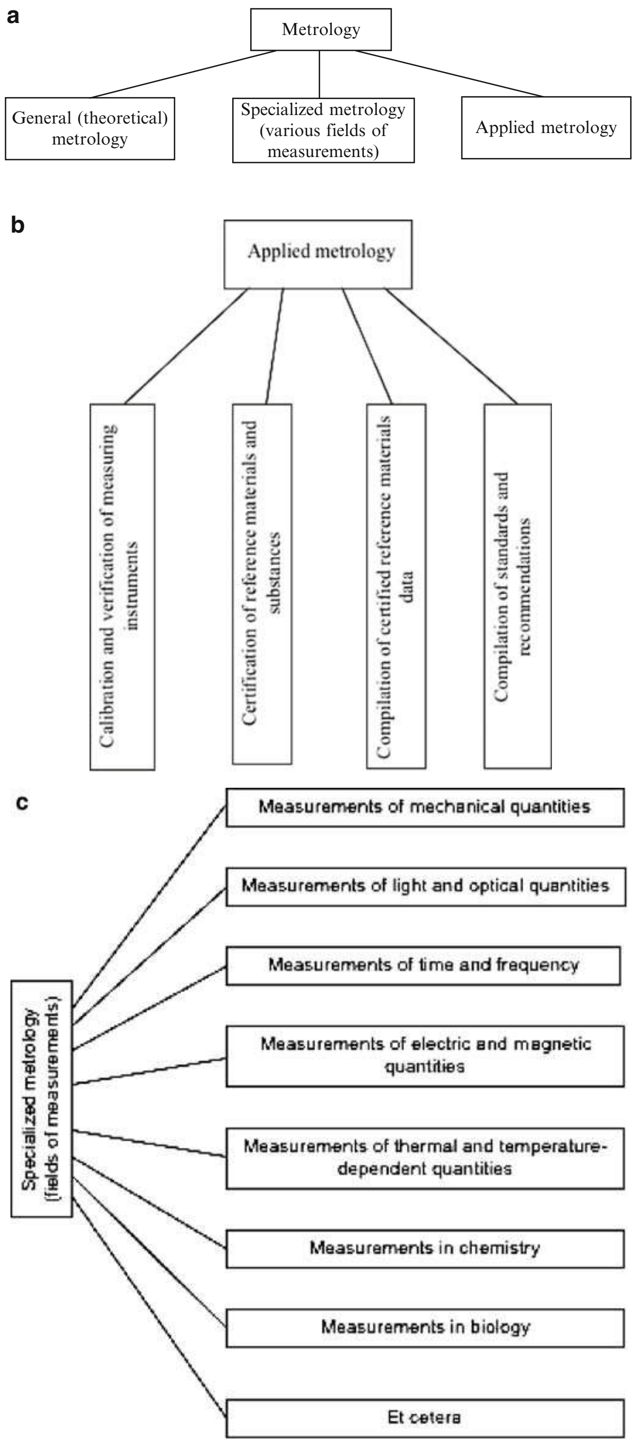


Fig. 1.1 Schematic picture of the basic problems of metrology: (**a**) metrology, (**b**) applied metrology, (**c**) specialized metrology, and (**d**) general metrology

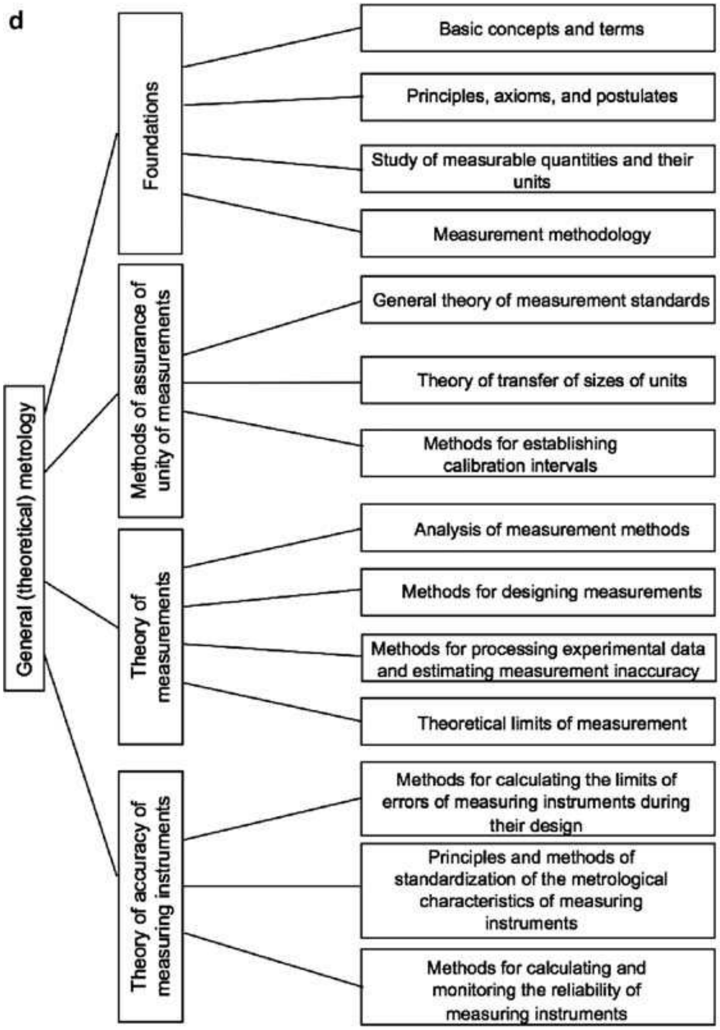


Fig. 1.1 (continued)

in any of these sciences, and already for this reason, they must be included among the problems addressed in metrology. An important result in this direction was the creation of the International System of Units SI.

2. General Theory of Measurement Standards

The units of quantities are reproduced with the help of *primary measurement standards*, which play an exceptionally important role in supporting the unity of measurements. The measurement standard of each unit is physically created based on the laws of specific fields of science and technology. Therefore, general metrology cannot answer the question of how a measurement standard should be constructed. But metrology must determine the criteria when a measurement standard must be created and how it should be maintained and used. It must also study the theory and methods of comparing measurement standards and monitoring their stability, as well as methods for expressing their inaccuracy. Practice raises many such purely metrological questions.

3. Theory of Transfer of the Sizes of Units into Measurement Practice

In order for the results of all measurements to be expressed in established units, all means of measurement (measures, instruments, measuring transducers, measuring systems) must be calibrated with respect to primary measurement standards. However, it is obviously infeasible to calibrate all these devices against primary standards directly. This problem is solved with the help of a system of *secondary* measurement standards, i.e., standards that are calibrated with respect to the primary standard, and *working* measurement standards, i.e., standards that are calibrated with respect to secondary standards. Thus the system of measurement standards has a hierarchical structure. The entire procedure of calibrating measurement standards and, with their help, the measuring instruments is referred to as transfer of the sizes of units into measurement practice. The final stages of transferring the sizes of units consist of calibration of the scales of the measuring instruments, adjustment of measures, and determination of the actual values of the quantities that are reproduced by them, after which all measuring instruments are checked at the time they are issued and then periodically during use.

The procedures involved in the transfer of the size of units into measurement practice raise a number of questions. For example, how many gradations of accuracy of measurement standards are required? How many secondary and working standards are required for each level of accuracy? How does the inaccuracy increase when the size of a unit is transferred from one measurement standard to another? How does this inaccuracy increase during the transfer from a measurement standard to a working measuring instrument? What should be the relation between the accuracy of a measurement standard and a measuring instrument being calibrated (verified) with respect to this standard? How should complex measurement systems be checked? Metrology should answer these questions.

The other blocks in the diagram of Fig. 1.1d do not require any explanations. We shall now turn to Fig. 1.1a.

Specialized metrology is comprised from specific fields of measurement. Examples of fields of measurements include linear–angular measurements, measurements of mechanical quantities, measurements of electric and magnetic quantities, and so on. The central problem arising in each field of measurement is the problem of creating conditions under which the measurements of the corresponding quantities are unified. For this purpose, in each field of measurement, a system of measurement standards is created, and methods for calibrating and checking the working measuring instruments are developed. The specific nature of each field of measurement engenders many problems characteristic of it. These problems are the domain of specialized metrology. However, there also arise many problems that are common to several fields of measurement. The analysis of such common problems and the development of methods for solving them belong to *general metrology*.

Applied metrology incorporates the metrological service and legislative metrology, and it is of great importance for achieving the final goals of metrology as a science. The metrological service checks and calibrates measuring instruments and certifies reference materials; in other words, it maintains the uniformity of measuring instruments employed in the country. The functions of legislative metrology are to enact laws that would guarantee uniformity of measuring instruments and unity of measurements. One aspect of legislative metrology concerns the system of physical quantities and the units to be employed uniformly across a country, which can only be established by means of legislation. Another aspect legislates the rules giving the right to manufacture measuring instruments and to check the state of these instruments when they are in use.

This is a good point at which to discuss the development of measurement standards. A measurement standard is always a particular measuring device: a measure, instrument, or measuring system. Such measuring devices were initially employed as measurement standards arbitrarily by simple volition of the institution responsible for correctness of measurements in the country. However, there is always the danger that a measurement standard will be destroyed, which can happen because of a natural disaster, fire, and so on. An arbitrarily established measurement standard, which is referred to as a *prototype measurement standard*, cannot be reproduced.

As a result, scientists have for a long time strived to define units of measurement so that the primary measurement standards embodying them could be reproducible. For this, the units of the quantities were defined based on natural phenomena. Thus, the *second* was defined based on the period of revolution of the Earth around the sun; the *meter* was defined based on the length of the Parisian meridian, and so on. Scientists hoped that these units would serve "for all time and for all peoples." Historically, this stage of development of metrology coincided with the creation of the metric system.

Further investigations revealed, however, that the chosen natural phenomena are not sufficiently unique or are not stable enough. This, however, did not undermine the idea to define units based on natural phenomena. It was only necessary to seek other natural phenomena corresponding to a higher level of knowledge of nature.

It was found that the most stable or even absolutely stable phenomena are characteristic of phenomena studied in quantum physics; it was further found that the physical constants can be employed successfully for the purpose of defining units and the corresponding effects can be employed for realizing measurement standards. The meter, the second, the ohm, and the volt have now been defined in this manner.

Based on achievements in quantum physics, the second is reproduced now by the cesium atomic standard. According to NIST, it is so accurate that it takes almost 20 million years to accumulate the drift of 1 s. One needs to only recall that when the distance between two markings on a platinum–iridium rod was adopted for the meter, for the most accurate measurement of length, the inaccuracy was not less than 10^{-6}. When the meter was later defined as a definite number (1,650,763.73) of wavelengths of krypton-86 radiation in vacuum, this inaccuracy was reduced to 10^{-7}–10^{-8}. Today, the definition of the meter is based on the velocity of light in vacuum, which now considered as exactly known physical constant. As a result, the inaccuracy in measuring length has been reduced by another order of magnitude (and can be reduced even more). Since 1990, the primary standard of the volt has been based on the Josephson constant and quantum Josephson effect. Its inaccuracy, expressed as one standard deviation, is 0.6 ppm. From the same time, the primary standard of the ohm has been based on the Von Klitsing constant and quantum Hall effect. Its inaccuracy is 0.2 ppm (one standard deviation). The accuracy of the standards of volt and ohm can further increase with the improvements in the accuracy of measuring the constants mentioned above.

It is interesting to consider the situation with the standard of ampere – one of the base units in SI. Its definition is based on the force between two wires through the current flows. It is unknown how to reproduce this unit according to this definition with sufficient accuracy. For example, NIST has achieved reproducing ampere in this way only with the standard deviation of 15 ppm, and even this accuracy can be maintained for 5 min. At the same time, ampere can obviously be reproduced using Ohm's law, from the standards of volt and ohm, thus obtaining the accuracy of around 0.7 ppm. In other words, one can create a standard of ampere that would be 20 times more accurate than what is possible through the absolute method (using direct measurements) according to its definition. In other words, the primary standard of ampere became unnecessary for measurements! Note that ampere still remains a base unit of system SI and it is still needed for dimensional equations.

The numerical values of the basic physical constants are widely used in various calculations, and therefore, these values must be in concordance with each other. To this end, all values of fundamental physical constants obtained by experiments must be adjusted. The most recent adjustment was carried out in 2002 and the results were published in 2005 [40].

As one can see from the problems with which metrology is concerned, it is an applied science. However, the subject of metrology – measurement – is a tool of both fundamental sciences (physics, chemistry, and biology) and applied disciplines, and it is widely employed in all spheres of industry, commerce, and in everyday life. No other applied science has such a wide range of applications, as does metrology.

We shall return once again to specialized metrology. A simple list of the fields of measurement shows that the measurable quantities and therefore measurement methods and measuring instruments are extremely diverse. What then do the different fields of measurement have in common? They are united by general or theoretical metrology and, primarily, the general methodology of measurement, methods for processing measurement data, and evaluating the inaccuracy of measurements. For this reason, the development of these branches of metrology is important for all fields of science and for all spheres of industry that employ measurements. The importance of these branches of metrology is also indicated by the fact that a specialist in one field of measurement can easily adapt to and work in a different field of measurement.

1.3 New Forms of International Cooperation in Metrology

Modern development of metrology is driven, on one hand, by the ever-increasing role of measurements in chemistry, biology, laboratory medicine, food production, environmental protection, and monitoring, with ever-higher requirements for accuracy and, on the other hand, with the expansion of international trade and industry globalization.

The accelerated development of international trade began with the emergence of the European Union (EU), which resulted in the tariff-free trade zone encompassing all its member countries. Then other regional trade agreements, such as North American Free Trade Agreement (NAFTA), appeared, targeting the removal of barriers in international trade.

Besides international trade, another trend in modern economy is globalization of industrial production. It is now common that a factory producing a certain product is situated in one country but uses components from suppliers in other countries, has research and development divisions yet in other countries, and maintains corporate and administrative services still elsewhere.

This expansion of international cooperation dramatically increased the demand for metrology and metrological services. It became obvious that the international unity of measurements, i.e., when measurements of the same quantities in different countries would agree with each other, can bring enormous cost savings. Just considering trade, Kaarls [30] notes that "…global trade in commodities amounts to more than 12 trillion USD, of which 80% affected by standards and regulation. The compliance costs are estimated to be about 10% of the product costs. The global markets of clinical chemistry and laboratory medicine and pharmaceuticals have a value of some 300 billion USD per year. Annual savings as a consequence of comparable, more accurate measurements results…will easily amount up to many billions of USD."

Alongside traditional measuring instruments, there emerged a tremendous internationally distributed bank of reference materials and substances. Their preparation and usage need to be regimented to ensure the unity of measurements in chemistry, laboratory medicine, and other areas with wide reliance on these materials. In principle, methods of solving these issues are similar to those in traditional areas of measurements, except for the extremely large number and variety of these materials.

The current stage of metrology development reflects the emergence of new international and regional metrological agreements. These agreements are especially important for developing nations, since every region usually includes at least one country with a well-established metrological service and a modern metrological scientific center.

New agreements can be divided into general and targeted. The former include EUROMET (European Collaboration in Measurement Standards) and NORAMET (North and Central American Cooperation in Metrology). Among the latter, we should especially point out EURACHEM/CITAC. EURACHEM is a network of organizations in Europe having the objective of establishing a system for traceability of chemical measurements and the promotion of good quality practice, which was initially organized by the EU. Subsequently, in 1993, the Cooperation of International Traceability in Analytical Chemistry (CITAC) was created as an international addition to EURACHEM. Thus, EURACHEM/CITAC have the mission to improve traceability in chemical measurements made anywhere in the world; in other words, they aim at providing unity of chemical measurements on the global scale.

Several targeted agreements focus on bringing order to the process of assigning rights to various laboratories to carry out certain types of important measurements, that is, to regiment laboratory accreditations. These agreements include ILAC (International Laboratory Accreditation Cooperation) and APLAC (Asia – Pacific Laboratory Accreditation Cooperation). The work on regimenting laboratory accreditation is being carried out under the slogan "Measured or tasted once – everywhere accepted!"

Other targeted agreements have the goal of facilitating the cooperation between laboratories engaged in measurements in different countries, resolving disputes, etc. When necessary, the laboratories establish working groups, which focus on specific issues and issue clarifications of methodological and terminological nature. But the most important role of regional bodies is the establishment of the procedure for the comparison of standards of the member countries. These regional comparisons avoid the direct comparison of national standards of all countries that joined the Metric Convention with international standards in BIPM, which would be physically impossible.

In addition to government-level agreements, successful nongovernment organizations in developed countries are also expanding their international cooperation. For example, National Conference of Standard Laboratories, which used to be a US organization, became international (NCSLI).

Many of these organizations often face common problems, and they form joint working groups to address them. CIPM provides support to these groups, and in turn, members of these groups often serve as members of CIPM's Consultative Committees. We should also mention that BIPM organized a Joint Committee for Guides in Metrology (JCGM), with BIPM's Director serving as the Chair of the Joint Committee. This committee has two working groups whose tasks include the improvement of terminology and the development and advocating of the Guide to the Expression of Uncertainty in Measurement (GUM) [2].

GUM represents the first recommendation for the estimation of inaccuracy of measurements developed under the auspices of BIPM. Such a recommendation had been long overdue and the need for is obvious: a uniform solution to this problem is necessary to correlate different measurement results regardless of where and when they were obtained. Consequently, this recommendation found an enthusiastic acceptance by the metrological community and became an unofficial international standard. It turned out, however, that the recommendation had a number of drawbacks [13, 16, 31, 42, 44], and Working Group 1 of JCGM set out in 2006 to prepare its new edition.

In summary, the activities described above indicate vigorous development of metrology and metrological service at the present time. The role of metrology in the modern society was the subject of an extensive report by Dr. Quinn, Director of BIPM, titled "Metrology, Its Role in Today's World." This report was included as the introductory chapter of monograph [36].

1.4 Postulates of the Theory of Measurements

Measurements are so common and intuitively understandable that one would think there is no need to identify the foundations on which measurements are based. However, a clear understanding of the starting premises is necessary for the development of any science, and for this reason, it is desirable to examine the postulates of the theory of measurements.

When some quantity characterizing a specific object is being measured, this object is made to interact with a measuring instrument. Thus, to measure the diameter of a rod, the rod is squeezed between the jaws of a vernier caliper; to measure the voltage of an electric circuit, a voltmeter is connected to it; and so on. The reading of the measuring instrument – the sliding calipers, voltmeter, and so on – gives an estimate of the measurable quantity, i.e., the result of the measurement. When necessary, the number of divisions read on the instrument scale is multiplied by a certain factor. In many cases, the result of measurement is found by a mathematical analysis of the indications of an instrument or several instruments. For example, the density of solid bodies, the temperature coefficients of the electric resistance of resistors, and many other physical quantities are measured in this manner.

The imperfection of measuring instruments, the inaccuracy with which the sizes of the units are transferred to them, as well as some other factors that we shall study below cause measurement errors. Measurement errors are in principle unavoidable, because a measurement is an experimental procedure and the true value of the measurable quantity is an abstract concept. As the measurement methods and measuring instruments improve, however, measurement errors decrease.

The introduction of measurable quantities and the establishment of their units lay at the foundation of measurements. Any measurement, however, is always performed on a specific object, and the general definition of the measurable quantity must be formulated taking into account the properties of the object and the objective of the measurement. The true value of the measurable quantity is essentially introduced and defined in this manner. Unfortunately, this important preparatory stage of measurements is usually not formulated.

To clarify this question, let us consider a simple measurement problem – the measurement of the diameter of a disk. First, we shall formulate the problem. The fact that the diameter of a disk is to be measured means that the disk, i.e., the object of study, is a circle. We note that the concepts "circle" and "diameter of a circle" are mathematical, i.e., abstract, concepts. The circle is a representation or model of the given body. The diameter of the circle is the parameter of the model and is a mathematically rigorous definition of the measurable quantity. Now, in accordance with the general definition of the true value of the measurable quantity, it can be stated that the true value of the diameter of the disk is the value of the parameter of the model (diameter of the disk) that reflects quantitatively the property of the object of interest to us; the ideal qualitative correspondence must be predetermined by the model.

Let us return to our example. The intended usage of the disk predetermines the permissible measurement error and the choice of an appropriate measuring instrument. By bringing the object into contact with the measuring instrument, we perform the measurement and obtain the measurement result. But the diameter of the circle is, by definition, invariant under rotation. For this reason, the measurement must be performed in several places. If the difference between the results of these measurements is less than the permissible measurement error, then any of the obtained results can be taken as the result of measurement. After the value of the measurable quantity, a concrete number, which is an estimate of the true value of the measurand, has been found, the measurement can be regarded as being completed.

But it may happen that the difference among the measurements in different places exceeds the permissible error. In this situation, we must conclude that within the required measurement accuracy, our disk does not have a unique diameter, as does a circle. Therefore, no concrete number can be taken, with prescribed accuracy, as an estimate of the true value of the measurable quantity. Hence, the adopted model does not correspond to the properties of the real object, and the measurement problem has not been correctly formulated.

If the object is a manufactured article and the model is a drawing of the article (including all the dimensions and tolerances), then any disparity between them means that the article is defective. If, however, the object is a natural object, then the disparity means that the model is not applicable and it must be reexamined.

Of course, even when measurement of the diameter of the disk is assumed to be possible, in reality, the diameter of the disk is not absolutely identical in different directions. But as long as this inconstancy is negligibly small, we can assume that the circle as a model corresponds to the object and therefore a constant, fixed true value of the measurable quantity exists, and an estimate of the quantity can be found as a result of measurement. Moreover, if the measurement has been performed, we can assume that the true value of the measurand lies somewhere near the obtained estimate and differs from it by not more than the limits of the measurement error.

Thus the idealization necessary for constructing a model gives rise to an unavoidable discrepancy between the parameter of the model and the real property of the object. We shall call this discrepancy the *threshold discrepancy*.

As we saw above, the error caused by the threshold discrepancy between the model and the object must be less than the total measurement error. If, however, this component of the error exceeds the limit of permissible measurement error, then it is impossible to make a measurement with the required accuracy. This result indicates that the model is inadequate. To continue the experiment, if this is permissible for the objective of the measurement, the model must be redefined. Thus, in the example of the measurement of the diameter of a disk, a different model could be a circle circumscribing the disk.

Another example, the measurement of the thickness of a sheet of a material, is given in GUM (sections D.3.2 and D.3.4). Without additional clarifications, the problem statement assumes that the sheet has constant thickness. Then, the model of the object comprises two parallel planes, and the distance between them is the model parameter that defines the measurable quantity and its true value.

Now let us turn to the measurement. By choosing an appropriate measurement instrument and bringing it in contact with the object, we obtain the value of the measurand, i.e., the sheet thickness. To verify the appropriateness of the model, we need to repeat the measurement in several points of the sheet. If the difference between the readings turns out to be significant, that is, greater than the limits of permissible measurement error, then the assumed model or the chosen model parameter do not correspond to the properties of the object. Hence, the model or its parameter must be redefined. Depending on the intended use of the sheet, a new parameter could be the maximum thickness or the thickness in certain given points. In either case, the model remains the same but the model parameters are different. In the former case, the parameter is the maximum thickness, and in the latter case there are different parameters in each point. Thus, in the latter case, we must view thickness measurements in each point as separate measurements, each with its own true value.

Similar to the example of disk diameter, different results of measurement of the sheet thickness indicate a discrepancy between the model and the object and hence the need to reconsider the model and/or the definition of the true value. In fact, as we just saw, the new definition may introduce multiple true values and consequently replace a single measurement with several separate measurements. Moreover, the new definition may lead to the necessity to use different measurement instruments, for example, instruments with a reduced contact area in the sheet thickness scenario.

One important corollary from the above discussion is that the concept of the true value is necessary to understand the process of measurement. The above discussion also suggests that there is a single underlying true value in every measurement. We consider this to be a fundamental principle of measurement and include it into the postulates below. It also reflects a different understanding of the concept of the true value from VIM [1]. We will carefully examine the VIM position on the concept of true value in Sect. 9.2.

The above examples are simple, but they exhibit features present in any measurement, although these features are not always so easily and clearly perceived as when measuring lineal dimensions.

The foregoing considerations essentially reduce to three prerequisites of a measurement:

1. A model must be specified that corresponds to the object under study, and some parameter of the model must be defined to correspond to the measurand.
2. The model of the object must permit the assumption that during the time required to perform the measurement, the parameter of the model corresponding to the measurand is constant.
3. The error caused by the threshold discrepancy between the model and the object must be less than the permissible measurement error.

The above prerequisites do not include a basic assumption behind any measurement that the general definition of the measurable quantity (e.g., length, time, electrical resistance, or whatever quantity is being measured) has been already introduced, and the corresponding measurement standards exist. The issues of measurable quantity definitions and the availability of standards are not directly related to the problem of estimating measurement accuracy, and for this reason, they are not studied here. These issues are investigated in several works; we in particular refer the reader to the book by B.D. Ellis [24] and the work of K.P. Shirokov [50].

Generalizing all three prerequisites, we formulate the following principle of metrology:

> *A measurement of a measurable quantity of an object with a given accuracy can be performed only if it is possible to associate, with no less accuracy, a determinate parameter of the model with that measurable quantity.*

We note that the value of the parameter of the model of an object introduced in this manner is the true value of the measurable quantity.

The foregoing considerations are fundamental, and they can be represented in the form of postulates of the theory of measurement [46], [52]:

(α) *The true value of the measurable quantity exists.*
(β) *There is a single true value in each measurement.*
(γ) *The true value of the measurable quantity is constant.*
(δ) *The true value cannot be found.*

The threshold discrepancy between the model and the object was employed above as a justification of the postulate (δ). However, other unavoidable restrictions also exist on the approximation of the true value of a measurable quantity. For example, the accuracy of measuring instruments is unavoidably limited. For this reason, it is possible to formulate the simple statement: *The result of any measurement always contains an error.*

We shall now discuss some examples of models that are employed for specific measurement problems.

Example 1.1. Measurement of the Parameters of Alternating Current

The object of study is an alternating current. The model of the object is a sinusoid

$$i = I_{\mathrm{m}} \sin(\omega t + \varphi),$$

where t is the time and I_{m}, ω, and φ are the amplitude, the angular frequency, and the initial phase, and they are the parameters of the model.

Each parameter of the model corresponds to some real property of the object and can be a measurable quantity. But, in addition to these quantities, several other parameters that are functionally related with them are also introduced. These additional parameters can also be measurable quantities. Some parameters can be introduced in a manner such that by definition they are not related with the “details” of the phenomenon. An example of such a parameter is the effective current

$$I = \sqrt{\frac{1}{T}\int_0^T i^2 dt},$$

where $T = 2\pi/\omega$ is the period of the sinusoid.

A nonsinusoidal current is also characterized by an effective current. However, in designing measuring instruments and describing their properties, the form of the current, i.e., the model of the object of study must be taken into account.

The discrepancy between the model and the object in this case is expressed as a discrepancy between the sinusoid and the curve of the time dependence of the current. In this case, however, only rarely it is possible to discover the discrepancy between the model and the object under study by means of simple repetition of measurements of some parameters. For this reason, the correspondence between the model and the object is checked differently, for example, by measuring the form distortion factor. If the discrepancy is detected, the model is usually redefined by replacing the sinusoid with a sum of a certain number of sinusoids.

Example 1.2. Measurement of the Parameters of Random Processes

The object of the study is some randomly changing quantity. The usual model is a stationary ergodic random process on the time interval T. The constant parameters of the process are the mathematical expectation $E[X]$ and the variance $V[X]$. Suppose that we are interested in $E[X]$. The value of this parameter in the mathematical model of the process is the true value of the measurand in this case. It can be estimated, for example, with the help of the formula

$$\overline{x} = \left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)_T,$$

where T is the observational time interval, x_i are the estimates of the realizations of the random quantity, whose variation in time forms a random process at times $t_i \in T$, and n is the total number of realizations obtained.

Repeated measurements on other realizations of the process can give somewhat different values of $\overline{x}$. The adopted model can be regarded as corresponding to the physical phenomenon under study, if the differences between the obtained estimates of the mathematical expectation of the process are much smaller than the permissible measurement error. If, however, these differences are close to the error or exceed it, then the model must be redefined, which is most simply done by increasing the observational interval T.

It is interesting to note that the definitions of some parameters seem, at first glance, to permit arbitrary measurement accuracy (if the errors of the measuring instrument are ignored). Examples of such parameters are the parameters of stationary random processes, the parameters of distributions of random quantities, and the average value of the quantity. One would think that to achieve the required accuracy in these cases, it is sufficient to increase the number of observations when performing the measurements. In reality, however, the accuracy of measurement is always limited, and in particular, it is limited by the correspondence between the model and the phenomenon, i.e., by the possibility of assuming that to the given phenomenon, there corresponds a stationary random process or a random quantity with a known distribution.

When a true value cannot be defined, then a measurement is impossible. For example, in the last few years, much has been written about measurements of variable and random quantities. However, these quantities, as such, do not have a true value, and for this reason, they cannot be measured.

For a random quantity, it is possible to measure the parameters of its distribution function, which are not random; it is also possible to measure the realization of a random quantity. For a variable quantity, it is possible to measure its parameters that are not variable; it is also possible to measure the instantaneous values of a variable quantity.

We shall now discuss in somewhat greater detail the measurement of instantaneous values of quantities. Suppose that we are studying an alternating current, the model of which is a sinusoid with amplitude I_m, angular frequency ω, and initial phase φ. At time t_1, there is an instantaneous value in the model, $i_1 = I_m \sin(\omega t_1 + \varphi)$, which corresponds to an instantaneous current. At a different time, there will be a different instantaneous value, but at each moment, it has some definite value.

Thus, there always exists a fixed parameter of the model corresponding to the measurable property of the object.

Measurement, however, is not instantaneous. The measurable quantity (the current in the above example) will change while the measurement is taken, and this will generate a specific error of the given measurement. The objective of the measurement determines a permissible level that the measurement error, including its component caused by the change in the measurable quantity during the measurement time, must not exceed. If this condition is satisfied, then the effect of the measurement time can be neglected, and one can assume to have obtained an estimate of the measured instantaneous current, i.e., the current strength at a given moment in time. In the literature, the expressions "measurement of a variable quantity" and "measurement of a random quantity" often refer to, respectively, measurement of instantaneous values and measurement of a realization of a random quantity. Such usage of these expressions is obviously incorrect.

Measurable quantities are divided into active and passive. Active quantities are quantities that can generate measurement signals without any auxiliary sources of energy; i.e., they act on the measuring instruments. Such quantities are the EMF, the strength of an electric current, mechanical force, and so on. Passive quantities cannot act on measuring instruments, and for measurements, they must be activated. Examples of passive quantities include mass, inductance, and electric resistance. Mass is usually measured based on the fact that in a gravitational field, a force proportional to the mass acts on the body. Electric resistance is activated by passing an electric current through a resistor. When measuring a passive quantity of an object, the object model is constructed for the active quantity (or quantities) that arises from the activation of passive quantities.

1.5 Classification of Measurements

In metrology there has been a long-standing tradition to distinguish direct, indirect, and combined measurements. In the last few years, metrologists have begun to divide combined measurements into strictly combined measurements and simultaneous measurements [12].

Direct measurements are measurements in which the object of study is made to interact with the measuring instrument, and the value of the measurand is read from the indications of the latter. Sometimes the instrumental readings are multiplied by some factor or adjusted by applying certain corrections.

In the case of indirect measurements, the value of the measurable quantity is found based on a known functional dependence between this quantity and other quantities called *arguments*. The arguments are found by means of direct and sometimes indirect measurements, and the value of the measurand is calculated according to the known dependence. For example, the density of a homogeneous solid body is found as the ratio of the mass of the body to its volume. To obtain the density, the mass, and volume of the body – the arguments – are measured directly, and the density is then computed from their measured values.

Sometimes direct and indirect measurements are not easily distinguished.

For example, an AC wattmeter has four terminals. The voltage applied to the load is connected to one pair of terminals, whereas the other pair of terminals is connected in series with the load. As is well known, the indications of a wattmeter are proportional to the power consumed by the load. However, the wattmeter does not respond directly to the measured power and its operation is based on the transformation of the strengths of two electric currents into a mechanical rotation. Given the principle of operation of the instrument, measurement of power by a wattmeter should be regarded as indirect.

In our case, it is important, however, that the value of the measurable quantity can be read directly from the instrument (in this case, the wattmeter). In this sense, a wattmeter is in no way different from an ammeter. For this reason, in this book, it is not necessary to distinguish measurement of power by a wattmeter and measurement of the strength of current by an ammeter: We shall categorize both cases as direct measurements. In other words, when considering a specific measurement as belonging to one or another category, we will ignore the internals of the measuring instrument employed.

A similar confusion may arise in the case of measurements performed with a measuring system or a chain of measuring instruments. A simple example of such measurements is the measurement of temperature with thermocouple and millivoltmeter. The thermocouple produces for each temperature the corresponding electromotive force (EMF) and the voltmeter measures this EMF. From the indication of the millivoltmeter and knowing the characteristics of the thermocouple, one can determine the temperature being measured.

The last instrument in the chain from which the measurement result is read (the millivoltmeter in our example) may be graduated directly in units of the measurand (the temperature) or in other units (for instance, one could just use a general purpose millivoltmeter in our example). In the former case, we would like to stress that the entire chain should be viewed as a single (albeit complex) instrument, and it should be calibrated as such. In particular, its intrinsic and additional errors should be rated for the entire unit. Inaccuracy of the measurements in this case is estimated using the methods for measurements with a single instrument as described in Chap. 4. In the latter case, that is, if the last measuring instrument is graduated in different units, this becomes an indirect measurement, and its inaccuracy is estimated according to the methods presented in Chap. 5.

Simultaneous and combined measurements are rather similar types of measurements. In both cases, their distinguishing property is that the objective of the measurement is to obtain values of several quantities rather than a single quantity as with direct and indirect measurements. Also, in both cases, measurable quantities are found by solving a system of equations, whose coefficients and certain terms are obtained as a result of measurements. Finally, in both cases, the method of least squares (see Chap. 6) is usually employed. But the difference is that in the case of combined measurements, several quantities of the same kind are measured, whereas in the case of simultaneous measurements, quantities of different kinds are measured at the same time. For example, a measurement, in which both the electric resistance of a resistor at temperature $+20°C$ and its temperature coefficient are found using the direct measurements of the resistance and temperature performed at different temperatures, is a simultaneous measurement. A measurement, in which the masses of separate weights in a set are found based on the known mass of one of them and by comparing with it the masses of different combinations of weights from the same set, is a combined measurement.

Depending on the properties of the object of study, the model adopted for the object, the definition of the measurable quantity given in the model, as well as on the method of measurement and the properties of the measuring instruments, the measurements in each of the categories mentioned above are performed either with single or with repeated observations. The method employed for processing the experimental data depends on the number of observations – are many measurements required or are one or two observations sufficient? If a measurement is performed with repeated observations, then, to obtain the result, the observations must be analyzed statistically. On the other hand, statistical methods are not required in the case of measurements with single observations. For this reason, we argue that the number of observations is an important classification criterion.

We shall term measurements performed with single observations as *single measurements* and measurements performed with repeated observations as *multiple measurements*. These terms have a natural intuitive meaning in direct measurements but need clarification for indirect measurements. An indirect measurement, in which the value of each of the arguments is found as a result of a single measurement, must be regarded as a single measurement. If, on the other hand, the values of the arguments were obtained by multiple measurements, the whole indirect measurement is considered a multiple measurement.

Measurements are also divided into static and dynamic measurements. Adhering to the concept presented in [51], we shall classify as static those measurements in which the measuring instruments are employed in the static regime and as dynamic those measurements in which the measuring instruments are employed in the dynamic regime. The static regime of a measuring instrument is a regime in which the output signal of the instrument can be regarded as constant. For example, for an indicating instrument, the regime is static if the signal is constant for a time sufficient to take the reading. A dynamic regime is a regime in which the output signal changes in time, so that to obtain a result or to estimate its accuracy, this change must be taken into account.

According to these definitions, static measurements include, aside from trivial measurements of length, mass, and so on, direct measurements of the average and effective (mean-square) values of alternating current by indicating instruments. A typical example of dynamic measurements is tracking the value of a quantity as a function of time by a recording instrument. Note that one can view such measurement as an infinite set of single instantaneous measurements; in this case, each instantaneous measurement would be considered static. Other examples of dynamic measurements are measurement of the magnetic flux by the ballistic method and measurement of the high temperature of an object based on the initial portion of the transfer function of a thermocouple put into contact with the object for a short time (the thermocouple would be destroyed if the contact time was long).

Static measurements also include measurements performed using digital indicating instruments. According to the definition of static measurements, for a measurement to be considered static, it is not important that the state of the elements in the device changes during the measurement. The measurement will also remain static when the indications of the instrument change from time to time, but each indication remains constant for a period of time sufficient for the indication to be read or recorded automatically.

A characteristic property of dynamic measurements is that to obtain results and estimate their accuracy in such measurements, it is necessary to know a complete dynamic characteristic of the measuring instrument: a differential equation, transfer function, and so on. (The dynamic characteristics of measuring instruments will be examined in Chap. 2.)

The classification of measurements as static and dynamic is justified by the difference in the methods employed to process the experimental data. At the present time, however, dynamic measurements as a branch of metrology are still in the formative stage.

The most important characteristic of the quality of a measurement is accuracy. The material base, which ensures the accuracy of numerous measurements performed in the economy, consists of measurement standards. The accuracy of any particular measurement is determined by the accuracy of the measuring instruments employed, the method of measurement employed, and sometimes by the skill of the experimenter. However, as the true value of a measurable quantity is always unknown, the errors of measurements must be estimated computationally. This problem is solved by different methods and with different accuracy.

In connection with the estimation of measurement accuracy, we shall distinguish measurements whose accuracy (or, more commonly, inaccuracy) is estimated before and after the measurement. We shall refer to them as measurements with a priori estimation of inaccuracy and measurements with a posteriori estimation of inaccuracy.

Measurements with a priori inaccuracy estimation must be performed according to an established procedure. Measurements of this type include all mass measurements.

Mass measurements (also called industrial measurements in [1]) are common. Their accuracy is predetermined by the types (brands) of measuring instruments indicated in the procedure, the techniques for using them, as well as the stipulated conditions under which the measurements are to be performed. Note that, in mass measurements, procedure for the a priori inaccuracy estimation is implicitly reflected in the overall measurement procedure: the person performing the measurement is interested only in the result of measurement, simply assuming that the accuracy will be adequate as long as he or she follows the procedure.

A posteriori estimation of inaccuracy is characteristic for measurements when it is important to know the accuracy of each result. We shall further divide measurements with a posteriori estimation of inaccuracy into two groups: measurements with universal estimation of inaccuracy and measurements with individual estimation of inaccuracy.

Measurements with universal estimation of inaccuracy are measurements in which the manufacturer specifications (rather than actual properties) of the measuring instruments employed are taken into account. These properties hold for all instruments of a given type; thus universal estimates remain valid when an instrument is replaced with another instrument of the same type.

Measurements with individual estimation of inaccuracy are measurements in which the inaccuracy estimation takes into account actual properties of the specific measuring instruments employed. These properties are usually established by calibration laboratories and are listed in calibration certificates.

In both cases, the conditions under which the measurements are performed are taken into account; this is done by obtaining and applying the influence quantities of the measurement conditions. In many cases, the influence quantities are measured; in other cases, they are estimated. We will refer to the measurements of influence quantities as *supplementary measurements*. Distinguishing supplementary measurements is useful for metrological purposes.

Here we would like to call attention to a fact whose validity and significance will become obvious from further discussion. Suppose that several measurements are performed using the same measuring instruments but with different methods of inaccuracy estimation. Although the same instruments are employed, these measurements will have different accuracy. The inaccuracy established by individual estimation will be less than the inaccuracy found by universal estimation.

The results of measurements with a priori and a posteriori inaccuracy estimation will be only rarely equally accurate. However, when measurements employ measuring instruments with different accuracy, the above conclusion will no longer be true. For example, measurement of voltage with a potentiometer of accuracy class 0.005, performed as a mass measurement, i.e., with a priori inaccuracy estimation, will be more accurate than measurement with an indicating voltmeter of class 0.5 and individual inaccuracy estimation.

Returning to the discussion of various measurement types, measurements are often performed during the preliminary study of a phenomenon. We shall call such measurements as *preliminary measurements*. The purpose of preliminary measurements is to determine the conditions under which some characteristic of the phenomenon can be observed repeatedly, so that its regular relations with other properties of the object, systems of objects, or with an external medium can be studied. As the objective of natural sciences is to establish and study regular relations between objects and phenomena, preliminary measurements are important in these fields. In particular, the first task of a scientist who is studying some phenomenon is usually to determine the conditions under which the phenomenon can be observed repeatedly in other laboratories and can be checked and confirmed.

Preliminary measurements are also required to construct a model of the object under study. For this reason, preliminary measurements are important in metrology as well.

Enormous literature exists on different aspects of measurements. As just one example, we can refer the reader to the book by Massey [38], which considered a number of these aspects.

1.6 Classification of Measurements Errors

Measurement accuracy is characterized by measurement error, limits of error, or uncertainty. A measurement of a quantity whose true value is A gives an estimate $\tilde{A}$ of that quantity. The absolute measurement error ζ expresses the difference between $\tilde{A}$ and A: $\zeta = \tilde{A} - A$. However, this equation cannot be used to find the error of a measurement for the simple reason that the true value of the measurable quantity is always unknown.

As mentioned previously, only in calibration of measuring instruments can one assume that the true value of the measurand is known, by taking the value of the measurement standard (often called “reference standard” in this context) as the true value of the measurand. Even then, strictly speaking, one finds the error of the device being calibrated and not of the measurement itself. The error of the measurement device found during calibration is called a *point estimate*.

In all other cases, the measurement accuracy is characterized by either limits of error or uncertainty, that is, by *intervallic estimates*. The calculation of these estimates is based on estimating errors contributed by various individual sources of inaccuracy; the latter are called *elementary errors* of the measurement.

The necessary components of any measurement are the method of measurement and the measuring instrument; in addition, measurements are often performed with the participation of a person. The imperfection of each component of measurement contributes to the measurement error. For this reason, in the general form,

$$\zeta = \zeta_{\mathrm{m}} + \zeta_{\mathrm{i}} + \zeta_{\mathrm{p}},$$

where ζ is the measurement error, ζ_{m} is the methodological error, ζ_{i} is the instrumental error, and ζ_{p} is the personal error.

Each component of the measurement error can in turn be caused by several factors. Thus, *methodological errors* can arise as a result of an inadequate theory of the phenomena on which the measurement is based and inaccuracy of the relations that are employed to find an estimate of the measurable quantity. In particular, the error caused by the threshold discrepancy between the model of a specific object and the object itself is a methodological error.

Instrumental measurement errors are caused by the imperfection of measuring instruments. Normally the intrinsic error of measuring instruments, i.e., the error obtained under reference conditions regarded as normal, is distinguished from additional errors, i.e., errors caused by the deviation of the influence quantities from their values under reference conditions. Properties of measuring instruments that cause the instrumental errors will be examined in detail in Chap. 2.

Human participants are responsible *for personal errors*. The individual characteristics of the person performing the measurement give rise to individual errors that are specific to that person. For example, in a measurement of high temperature using an optical pyrometer, a human must detect the moment when the image of a filament vanishes on the screen of the pyrometer. This moment (as detected) will depend on the person's perception. Another typical example includes incorrect reading of an instrument indication when it falls in-between graduation marks of the instrument scale.

Thanks to improvements in the reading and regulating mechanisms of measuring instruments, personal errors are usually insignificant for modern measuring instruments. In particular, they are virtually nonexistent for digital instruments.

The foregoing classification of measurement errors is based on the cause of the errors. Another important classification of measurement errors is based on their properties. In this respect, systematic and random errors are distinguished.

A measurement error is said to be *systematic* if it remains constant or changes in a regular fashion in repeated measurements of one and the same quantity. The observed and estimated systematic error is eliminated from measurements by introducing corrections. However, it is impossible to eliminate completely the systematic error in this manner. Some part of the error will remain and then this residual error will be the systematic component of the measurement error.

To define a random measurement error, imagine that some quantity is measured several times. If there are differences between the results of separate measurements and these differences cannot be predicted individually, then the error from this scatter of the results is called the *random error*.

The division of measurement errors into systematic and random is important, because these components are manifested differently and different approaches are required to estimate them. Random errors are discovered by performing measurements of one and the same quantity repeatedly under the same conditions, whereas systematic errors can be discovered experimentally either by comparing a given result with a measurement of the same quantity performed by a different method or by using a more accurate measuring instrument. However, systematic errors are normally estimated by theoretical analysis of the measurement conditions, together with the known properties of a measurand and of measuring instruments. Other specifics of the terms systematic and random errors are discussed in Sect. 4.2.

In speaking about errors, we shall also distinguish gross or outlying errors and blunders. We shall call an error *gross* or *outlying* if it significantly exceeds the error justified by the conditions of the measurements, the properties of the measuring instrument employed, the method of measurement, and the qualifications of the experimenter. Such measurements can arise, for example, as a result of a sharp, brief change in the grid voltage (if the grid voltage in principle affects the measurements).

Outlying or gross errors in multiple measurements are discovered by statistical methods and are usually eliminated from analysis.

Blunders occur as a result of errors made by the experimenter. Examples are a slip of the pen when writing up the results of observations, an incorrect reading of the indications of an instrument, and so on. Blunders are discovered by nonstatistical methods, and they must always be eliminated from the analysis.

Measurement errors are also divided into static and dynamic. Static errors are exhibited by static measurements. Dynamic errors are present in dynamic measurements and are caused by the inertial properties of measuring instruments. For example, if a varying quantity is recorded with the help of a recording instrument, then the difference between the obtained function and the actual quantity as it changes with time (taking into account the necessary scale transformations) is the dynamic error of the given dynamic measurement. In this case, the dynamic error is also a function of time, and the instantaneous dynamic error can be determined for each moment in time.

We shall now study the case when the process is recorded by measuring individual instantaneous values. It is clear that if within the time of a single measurement, the measurable quantity does not change significantly and the instantaneous values of the process are obtained at known times and sufficiently frequently, then the collection of points ultimately obtained gives an arbitrarily close approximation of the continuous recording. Thus, there will be no dynamic error here.

The inertial properties of an instrument can be such, however, that the changes in the measurable quantity during the time necessary to perform a point measurement will lead to a definite error in the measurements of the point values. In this case, the obtained collection of point values will deviate from the measurable quantity as it changes in time, and their difference, exactly as in the above case of a recording instrument, will give the dynamic error. It is natural to call the errors of separate point measurements as *instantaneous dynamic errors.*

1.7 General Approach to Evaluation of Measurement Inaccuracy

Measurements are regarded metrologically to be better the lower their inaccuracy is. However, measurements must be reproducible, because otherwise they lose their objective character and therefore become meaningless.

The reproducibility of a measurement depends on proper estimates of its inaccuracy. For example, consider a measurement of the length of a certain object. Assume an experimenter measures this length to be 3.000 m with proper limits of errors (as warranted by the measurement instruments and procedure) to be ± 0.3 cm. If the experimenter estimates the limits of error too conservatively to be ± 0.5 cm, then the accuracy of this measurement will be unnecessarily low, but it will be reproducible: it will be confirmed if someone else measures this length with higher accuracy. However, if the first experimenter erroneously estimates the limits of error to be ± 0.01 cm, this measurement will no longer be reproducible. A more accurate measurement will refute it.

Thus, correctly estimated measurement inaccuracy permits comparing the obtained result with the results obtained by other experimenters. The fact that the correctness of a given estimate is later confirmed in a more accurate measurement attests to the high skill of the experimenter. But the above argument exposes contradictory tendencies. On one hand, every experimenter wants to present his or her measurement as being as high quality as possible; on the other hand, the measurement result must be reproducible, and this suggests conservative estimation of the accuracy.

With regard to the above contradiction, we stress that while high quality of a measurement is desirable, the reproducibility (or, said differently, reliability) of the measurement is mandatory. Thus, it is better to err on the side of caution and be biased toward reliability, that is, conservative inaccuracy estimations. This conclusion should be considered as the following principle of the estimation of measurement inaccuracy:

The estimate of the inaccuracy of measurement must be an upper-bound estimate.

The inaccuracy estimation for any measurement result is based on the estimates of elementary errors of this measurement. Therefore, to satisfy the above principle, the estimates of the elementary errors must also be upper-bound estimates. At the same time, combining the elementary errors into the overall inaccuracy estimate of the measurement should be done without introducing unwarranted additional inaccuracy exaggeration, so that the overall inaccuracy estimate is only minimally exaggerated.

We should also stress that the correctness of an estimate of inaccuracy of a measurement cannot be checked based on data obtained in that same measurement. In any given measurement, all obtained experimental data and other reliable information, for example, corrections to the indications of instruments, are employed to find the measurement result, and the error must be estimated with additional information about the properties of the measuring instruments, the conditions of the measurements, and the theory. There is no point in performing a special experiment to check or estimate the measurement error or uncertainty. It would entail organizing in parallel with the given measurement a more accurate measurement of the same measurable quantity. Then the given measurement would be meaningless: Its result would be replaced by the result of the more accurate measurement. The problem of estimating the error in the given measurement would be replaced by the problem of estimating the error of the more accurate measurement; i.e., the basic problem would remain unsolved.

The correctness of estimates of errors and uncertainty is nonetheless checked. It is confirmed either by the successful use of the measurement result for the purpose intended or by the fact that the measurement agrees with the results obtained by other experimenters. As in the case of measurement of physical constants, the correctness of the estimates of uncertainties is sometimes checked with time as a result of improvements in measuring instruments.

1.8 Presentation of Measurement Results

If $\tilde{A}$ is the result of a measurement and Δ_{U} and Δ_{L} are the upper and lower limits of the error in the measurement, then the result of the measurement and the measurement inaccuracy can be written in the form

$$\tilde{A},\ \Delta_{\mathrm{U}},\ \Delta_{\mathrm{L}}.$$

For example, a measurement result and its inaccuracy could be represented as $\tilde{A} = 1.153\,\mathrm{cm}$, $\Delta_{\mathrm{U}} = +0.002\,\mathrm{cm}$, and $\Delta_{\mathrm{L}} = -0.001\,\mathrm{cm}$. Often, $|\Delta_{\mathrm{U}}| = |\Delta_{\mathrm{L}}| = \Delta$. Then, the result and the inaccuracy are written in the form $\tilde{A} \pm \Delta$.

But more often, the inaccuracy is expressed as uncertainty. In this case, the corresponding probability that the error is within the specified limits must be given. For uniformity, it is recommended that the probability be given in parentheses after the value of the uncertainty or a symbol of a measurand.

For example, if a measurement gives the value of the voltage, 2.62 V, and the uncertainty of this result, $u = \pm 2\%$, was calculated for the probability 0.95, then the result will be written in the form

$$\tilde{U} = 2.62\,\mathrm{V},\ \ u = \pm 2\%(0.95)$$

or, in the more compact form,

$$U_{0.95} = (2.62 \pm 0.05)\,\mathrm{V}.$$

The compactness remark refers to the method for indicating the probability and is unrelated to the fact that the uncertainty is given in the relative form in the first case and in the absolute form in the second case. If the confidence probability is not indicated in the measurement result, then the inaccuracy must be assumed to have been estimated without the use of probability methods. Although an inaccuracy estimate obtained without the use of probability methods can be reliable, it cannot be associated with any probability value. Thus, the probability should not be indicated. To repeat, in this case, we have the limits of error of a measurement rather than the uncertainty.

The above representations of inaccuracy are desirable for the final result, intended for direct practical application, for example, in quality control. In this case, it is usually convenient to express the total inaccuracy estimation. In many cases, however, it is desirable to know not the total inaccuracy estimation but the characteristics of the random and systematic components separately. Such a representation of the inaccuracy makes it easier to analyze and determine the reasons for any discrepancy between the results of measurements of the same quantity performed under different conditions. An analysis of this kind is usually necessary in the case of measurements performed for scientific purposes, for example, measurements of physical constants. It is also desirable to record the components separately in those cases when the result of a measurement is to be used for calculations together with other data that are not absolutely precise. For example, in indirect measurements, when the arguments are measured directly, separate recording of the random and systematic errors of the measurements of the arguments makes it possible to estimate more accurately the uncertainty of the result of the overall indirect measurement. We will see this in Chap. 5.

For scientific measurements, apart from the inaccuracy expressions given above, it is helpful to describe the basic sources of error together with an estimate of their contribution to the total measurement uncertainty. For a random error, it is of interest to present the form and parameters of the distribution function of the observations and how the distribution function was determined (the method employed for testing the hypothesis regarding the form of the distribution function, the significance level used in this testing, etc.).

The inaccuracy in the results of mass measurements is usually not indicated at all, because it is estimated beforehand, and the estimation is known prior to the measurement. In mass measurements, the number of significant digits in the result of a measurement reflects the accuracy of the measurement.

In other measurements, the inaccuracy must be estimated and expressed explicitly.

As measurement inaccuracy determines only the vagueness of the result, the inaccuracy need not be known precisely. For this reason, in its final form, the inaccuracy is customarily expressed with only one or two significant digits. Two digits are retained for the most accurate measurements and if the most significant digit of the number expressing the inaccuracy is less than 3. However, in intermediate calculations, depending on the computational operations performed, one or two significant digits more than will be needed for the result should be retained so that the rounding error would not accumulate and distort the result.

The numerical value of the measurement result must have the last decimal digit of the same rank as the last digit in its inaccuracy estimation. There is no point in including more digits, because this will not reduce the inaccuracy of the result. But fewer digits, which can result from further rounding off the number, would increase the inaccuracy thus artificially reducing the accuracy of the result below that provided by the measurement employed.

For example, if the result of the measurement is 85.6342 and the limits of error are ± 0.04, then the result should retain only four significant digits: 85.63. If the same result has limits of error ± 0.012, then it should be expressed as 85.634.

If the rules presented above are used, then the number of significant digits in the measurement result makes it possible to judge approximately the accuracy of a measurement: the inaccuracy can reach at most three units in the next-to-last digit of the result. Returning to the above example, if we only know the result of 85.634, we can tell that according to the rules, the worse inaccuracy could have been ± 0.029. Indeed, any higher inaccuracy would have caused one to retain fewer digits in the result.

When retaining a proper number of significant digits in observations and measurement results, one must round the numbers involved. The rounding should be done according to the following rules:

1. The last retained digit is not changed if the adjacent digit being discarded is less than 5. Discarded digits in the whole part of the number are replaced by 0's and dropped in decimal fraction part.

Examples. Rounding the number 32,453 to four significant digits results in the number 32,450. Rounding the number 165.245 to four significant digits results in the number 165.2.

2. The last digit retained is increased by 1 if the adjacent digit being discarded is greater than 5 or if it is equal to 5 and there are digits other than 0 to its right.

Examples. If three significant digits are retained, the number 18.598 is rounded to 18.6 and the number 152.56 is rounded to 153.

3. If the digit being discarded is equal to 5 and the digits to its right are unknown or are equal to 0, then the last retained digit is not changed if it is even and it is increased by 1 if it is odd.

Examples. If two significant digits are retained, the number 10.5 is rounded to 10 and the number 11.50 is rounded to 12.

4. If the decimal fraction in the numerical value of the result of a measurement terminates in 0's, then the 0's are dropped only up to the digit that corresponds to the rank of the least significant digit of the numerical value of the inaccuracy estimation.

The foregoing rules were established by convention, and for calculations performed by humans, they are entirely satisfactory. In the case of calculations performed with the help of computers, however, rounding depending on the evenness or oddness of the last retained digit [rule (3)] is inconvenient, because it complicates the algorithm. It has been suggested that this rule be dropped and the last retained figure not be changed, irrespective of whether it is even or odd. This suggestion, however, has not been adopted. The main objection is that such rounding, if applied consecutively to intermediate results, can significantly distort the final result.

We shall now estimate the relative rounding error, based on the observation that the limits of error caused by the rounding are equal to one-half the last digit in the numerical value of the result of the measurement. Assume, for example, that the measurement result is expressed as a number with two significant figures. Then the minimum number will be equal to 10 and the maximum number will be equal to 99. Therefore, the relative rounding error ε_2 of a result with two significant digits will be $0.5\% < \varepsilon_2 \leq 5\%$.

If the result of a measurement is expressed with three significant figures, this error will fall in the range $0.05\% < \varepsilon_3 \leq 0.5\%$, and so on. Thus, the limits of error obtained above show the effect of rounding off the result on the measurement error.

Chapter 2
Measuring Instruments and Their Properties

2.1 Types of Measuring Instruments

Measuring instruments are the technical objects that are specially developed for the purpose of measuring specific quantities. A general property of measuring instruments is that their accuracy is known. Measuring instruments are divided into material measures, measuring transducers, indicating instruments, recording instruments, and measuring systems.

A *material measure* is a measuring instrument that reproduces one or more known values of a given quantity. Examples of measures are balance weights, measuring resistors, measuring capacitors, and reference materials. Single-valued measures, multiple-valued measures, and collections of measures are distinguished. Examples of multiple-valued measures are graduated rulers, measuring tapes, resistance boxes, and so on. Multiple-valued measures are further divided into those that reproduce discrete values of the corresponding quantities, such as resistance boxes, and those that continuously reproduce quantities in some range, for example, a measuring capacitor with variable capacitance. Continuous measures are usually less accurate than discrete measures.

When measures are used to perform measurements, the measurands are compared with the known quantities reproduced by the measures. The comparison is made by different methods, but so-called *comparators* are a specific means that are used to compare quantities. A comparator is a measuring device that makes it possible to compare similar quantities and has a known sensitivity. The simplest comparator is the standard equal-armed pan balance.

In some cases, quantities are compared without comparators, by experimenters, with the help of their viewing or listening perceptions. For instance, when measuring the length of a body with the help of a ruler, the ruler is placed on the body and the observer fixes visually the graduations of the ruler (or fractions of a graduation) at the corresponding points of the body.

A *measuring transducer* is a measuring instrument that converts the measurement signals into a form suitable for transmission, processing, or storage. The measurement information at the output of a measuring transducer typically cannot be directly observed by the experimenter.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_2,

One must distinguish measuring transducers and the transforming elements of a complicated instrument. The former are measuring instruments, and as such, they have rated (i.e., listed in documentation) metrological properties (see below). The latter, on the other hand, do not have an independent metrological significance and cannot be used separately from the instrument of which they are a part.

Measuring transducers are diverse. Thermocouples, resistance thermometers, measuring shunts, and the measuring electrodes of pH meters are just a few examples of measuring transducers. Measuring current or voltage transformers and measuring amplifiers are also measuring transducers. This group of transducers is characterized by the fact that the signals at their inputs and outputs are a quantity of the same kind, and only the magnitude of the quantity changes. For this reason, these measuring transducers are called *scaling measuring transducers*.

Measuring transducers that convert an analog signal at the input into a discrete signal at the output are called analog-to-digital converters. Such converters are manufactured either as autonomous, i.e., independent measuring instruments, or as units built into other instruments, in particular, in the form of integrated microcircuits. Analog-to-digital converters are a necessary component of a variety of digital devices, but they are also employed in monitoring, regulating, and control systems.

An *indicating instrument* is a measuring instrument that is used to convert measurement signals into a form that can be directly perceived by the observer. Based on the design of the input circuits, indicating instruments are just as diverse as measuring transducers, and it is difficult to survey all of them. Moreover, such a review and even classification are more important for designing instruments than for describing their general properties.

A common feature of all indicating instruments is that they all have readout devices. If these devices are implemented in the form of a scale and an indicating needle, then the indications of the instrument are a continuous function of the magnitude of the measurable quantity. Such instruments are called analog instruments. If the indications of instruments are in a digital form, then such instruments are called digital instruments.

The above definition of digital instruments formally includes two types of devices. The first type, which includes automatic digital voltmeters, bridges, and similar instruments, performs all measuring transformations in a discrete form; in the second type, exemplified by induction meters for measuring electrical energy, all measuring transformations of signals occur in an analog form and only the output signal assumes a discrete form. The conversions of measurement information into a discrete form have several specific features. Therefore, only instruments in which the measurement conversions occur in a discrete form are usually considered to be digital instruments.

The indications of digital instruments can be easily recorded and are convenient for entering into a computer. In addition, their design usually makes it possible to obtain significantly higher accuracy than the accuracy of analog instruments. Moreover, when digital instruments are employed, no reading errors occur. However, with analog instruments, it is easier to judge trends in the variation of the measurands.

In addition to analog and digital instruments, there also exist analog-discrete measuring instruments. In these instruments, the measuring conversions are performed in an analog form, but the readout means are discrete (but not digital). Analog-discrete instruments combine the advantages of both analog and digital instruments. Mentioned above induction meters for measuring electric energy are examples of such hybrid instruments.

In many cases, measuring instruments are designed to record their indications. Such instruments are called *recording instruments.* Data can be recorded in the form of a continuous record of the variation of the measurand in time, or in the form of a series of discrete points. Instruments of the first type are called automatic-plotting instruments, and instruments of the second type are called printing instruments. Printing instruments can record the values of a measurand in digital form. Printing instruments give a discrete series of values of the measurand with some time interval. The continuous record provided by automatic-plotting instruments can be regarded as an infinite series of values of the measurand.

Sometimes measuring instruments are equipped with induction, photo-optical, or contact devices and relays for purposes of control or regulation. Such instruments are called regulating instruments. Regulating units typically lead to some reduction of the accuracy of the measuring instrument.

Measuring instruments also customarily include null indicators, whose primary purpose is to detect the presence of a nonzero signal. The reason for them to be considered measuring instruments is that a null indicator, such as a galvanometer, can often be used as a highly sensitive indicating instrument.

A *measuring system* is a collection of functionally integrated measuring, computing, and auxiliary devices connected to each other with communication channels.

2.2 Metrological Characteristics of Measuring Instruments

We shall divide all characteristics of measuring instruments into two groups: metrological, which are significant for using a measuring instrument in the manner intended, and secondary. We shall include in the latter such characteristics as mass, dimensions, and degree of protection from moisture and dust. We shall not discuss secondary characteristics because they are not directly related with the measurement accuracy, even though they sometimes influence the selection and application of an instrument.

By metrological characteristics of a measuring instrument, we mean the characteristics that make it possible to judge the suitability of the instrument for performing measurements in a known range with known accuracy. A simple example of a metrological characteristic common to all measuring instruments except single measures (i.e., measures reproducing a single value of a quantity) is the measurement range of the instrument. We will call metrological characteristics that are established before or during the design and development of the instrument as *nominal metrological characteristics*. Examples of such a characteristic are the nominal value of a measure (10 Ω, 1 kG, etc.), the measurement range of an instrument (0–300 V, 0–1,200°C, etc.), the conversion range of a transducer, the value of the scale factor of an instrument scale, and so on.

The relation between the input and the output signals of indicating instruments and transducers is determined by the transfer function. For indicating instruments, this relation is determined by the instrument scale, whereas for measuring transducers, it is determined by a graph or an equation. If this graph or equation had been determined and specified before the transducer was developed (or during its development), then the graph or equation represents a nominal metrological characteristic.

The real characteristics of measuring instruments differ from the nominal characteristics because of fabrication inaccuracies and changes occurring in the corresponding properties in time. These differences between nominal and real metrological characteristics lead to the error of the instrument.

Ideally, a measuring instrument would react only to the measured quantity or to the parameter of the input signal of interest, and its indication would not depend on the external conditions, such as the power supply regime, temperature, and so on. In reality, the external conditions do affect the indications of the instrument. The quantities characterizing the external conditions affecting the indications of a measuring instrument are called *influence quantities*.

For some types of measuring instruments, the dependence of the output signal or the indications on a given influence quantity can be represented as a functional dependence, called the *influence function*. The influence function can be expressed in the form of an equation (e.g., the temperature dependence of the EMF of standard cells) or a graph. In the case of a linear dependence, it is sufficient to give the coefficient of proportionality between the output quantity and the influence quantity. We call this coefficient the *influence coefficient*. Influence coefficients and functions make it possible to take into account the conditions under which measuring instruments are used, by introducing the corresponding corrections to the obtained results.

The imperfection of measuring instruments is also manifested because when the same quantity is measured repeatedly under identical conditions, the results can differ somewhat from one another. If these differences are significant, the indications are said to be nonrepeatable.

The inaccuracy of a measuring instrument is usually characterized by its error. Taking an indicating instrument as an example, let the true value of a quantity at the input of the instrument be A_t and the instrument indication be the value A_r. The absolute error of the instrument will be

$$\zeta = A_r - A_t.$$

If the indications of the repeated measurements of A_t are somewhat different, (but not enough to be considered nonrepeatable), one can talk about a random component of instrument error. For analog instruments, the random component of instrument error is normally caused by friction in the supports of a movable part of the instrument and/or by hysteresis phenomena. The limits of this error component can be found directly if the quantity measured by the instrument can be varied continuously, which is the case with, e.g., the electric current or voltage. The common method involves driving the indicator of the instrument continuously up to the same scale marker, once from below and once from above the marker. To compensate for friction (and/or hysteresis), the input signal that drives the indicator to the marker from below needs to be higher than what it would have been without friction; the input signal that drives the indicator to the same marker from above will be smaller. We will call *the dead band* the absolute value of the difference between the two values of the measurand that are obtained in such a test corresponding to a given scale marker of the instrument. The dead band gives the range of possible values of the random component of instrument error, and one half of this length is the limiting value of the random error.

There are also several instrument types, notably, weighing scales, whose indications cannot vary continuously. The random error of weighing scales is usually characterized by the standard deviation [7]. This characteristic of an instrument is calculated from the changes produced in the indications of the scales by a load with a known mass; the test is performed at several scale markers, including the limits of the measurement range. One method for performing the tests and the computational formula for calculating the standard deviation of weighing scales are presented in [7].

Measuring instruments are created to bring certainty into the phenomena studied and to establish regular relations between the phenomena. Thus, the uncertainty created by the nonrepeatability of instrument indications interferes with using an instrument in the manner intended. For this reason, the first problem that must be solved when developing a new measuring device is to make its random error insignificant, i.e., either negligibly small compared with other errors or falling within permissible limits of error for measuring devices of the given type. We should note here that because uncertainty of instrument indications represents only a random component of its inaccuracy, the term "uncertainty" cannot replace the term "limits of error" as applied to measuring instruments.

If the random error is insignificant and the elements determining instrument accuracy are stable, then by calibration, the measuring device can always be "tied" to a corresponding measurement standard and the potential accuracy of the instrument can be realized.

The value of the measurand corresponding to the interval between two neighboring markers on the instrument scale is called the *value of a scale division*. Similarly, the *value of the least significant digit* is the value of the measurand corresponding to one increment of the least significant digit of a digital readout device.

The *sensitivity* of a measuring instrument is the ratio of the change in the output value of the measuring instrument to the corresponding change in the input value of the quantity that causes the output value to change. The sensitivity can be a nominal metrological characteristic or an actual characteristic of a real instrument.

The *discrimination threshold* is the minimum change in the input signal that causes an appreciable change in the output signal.

The *resolution* is the smallest interval between two distinguishable neighboring discrete values of the output signal.

Instability (of a measuring instrument) is a general term that expresses the change in any property of the measuring instrument in time.

Drift is the change occurring in the output signal (always in the same direction) in the absence of the input signal over a period of time that is significantly longer than the time needed to perform a measurement with a given measuring instrument. The presence of drift entails the need to reset the zero indication of the instrument.

The drift and the instability do not depend on the input signal, but they can depend on the external conditions. The drift is usually determined in the absence of the signal at the input.

The metrological characteristics of measuring instruments should also include their dynamic characteristics. These characteristics reflect the inertial properties of measuring instruments. It is necessary to know them to correctly choose and use many types of measuring instruments. The dynamical characteristics are examined below in Sect. 2.5.

The properties of measuring instruments can normally be described based on the characteristics enumerated above. For specific types of measuring instruments, however, additional characteristics are often required. Thus, for the gauge rods, the flatness and degree of polish are important. For voltmeters, the input resistance is important. We shall not study such characteristics, because they refer only to individual types of measuring instruments.

2.3 Rating of the Errors of Measuring Instruments

Measuring instruments can only be used as intended when their metrological properties are known. In principle, the metrological properties can be established by two methods. One method is to find the actual characteristics of a specific instrument. In the second method, the nominal metrological characteristics and the permissible deviations of the real characteristics from the nominal characteristics are given.

The first method is laborious, and for this reason, it is used primarily for the most accurate and stable measuring instruments. Thus, the second method is the main method. The nominal characteristics and the permissible deviations from them are given in the technical documentation when measuring instruments are designed, which predetermines the properties of measuring instruments and ensures that they are interchangeable.

In the process of using measuring instruments, their real properties are checked to determine whether these properties deviate from the established nominal characteristics. If some real property deviates from its nominal value by an amount more than allowed, then the measuring instrument is adjusted, refurbished, or discarded and no longer used.

Thus, the choice of the nominal characteristics of measuring instruments and the designation of permissible deviations of the real characteristics from them –

rating of the metrological characteristics of measuring instruments – are of great importance for measurement practice. The practice of rating the metrological characteristics of measuring instruments has evolved over time, and we will examine it next.

Both the production of measuring instruments and the rating of their characteristics initially arose spontaneously in each country. Later, rules that brought order to the rating process were established in all countries with significant instrument production. The recommendations developed at this time by international organizations, primarily Publication 51 of the International Electrotechnical Commission (IEC) and a number of publications by the International Organization of Legal Metrology (OIML), were of great importance for standardizing the expression of rated characteristics [8, 9]. The terminological documents are also extremely valuable for developing rating procedures [1, 10, 12].

We shall now return to the gist of the problem. The values of nominal metrological characteristics, such as the upper limits of measurement ranges, the nominal values of the measures, the scale factors of instruments and so on, are chosen from a standardized series of values of these characteristics. A more difficult task is to rate the accuracy characteristics: errors and instability.

Despite the efforts of designers, the real characteristics of measuring instruments depend to some extent on the external conditions. For this reason, the conditions are determined under which the measuring instruments are to be calibrated and checked, including the nominal values of all influence quantities and the ranges of their permissible deviation from the nominal values. These conditions are called *reference conditions*. The error of measuring instruments under reference conditions is called the *intrinsic error.*

In addition to the reference conditions and intrinsic errors, the *rated operating conditions* of measuring instruments are also established, i.e., the conditions under which the characteristics of measuring instruments remain within certain limits and the measuring instruments can be employed as intended. Understandably, errors in the rated operating conditions are larger than errors under the reference conditions. This change is characterized by specifying the limits of the *additional error* (the additional error the instrument can have due to deviation of the corresponding influence quantity from the reference condition), the permissible value of the corresponding influence quantity, or by indicating the limits of the permissible error under the rated operating conditions (the overall possible error of the instrument).

The errors of measuring instruments are expressed not only in the form of absolute and relative errors, adopted for estimating measurement errors, but also in the form *of fiducial errors.* The fiducial error is the ratio of the permissible limits of the absolute error of the measuring instrument to some standardized value – *fiducial value*. The latter value is established by standards on separate types of measuring instruments; we discuss these rules later in this section. The fiducial error is somewhat similar to relative error but, since it is normalized to a constant standardized value, the fiducial error is constant across the entire measurement range of the device. The purpose of fiducial errors is that they make it possible to compare the accuracy of measuring instruments that have different measurement ranges. For example, the accuracy of an ammeter with a measurement limit of 1 A and permissible absolute error of 0.01 A has the same fiducial error of 1% (and in this sense, the same accuracy) as an ammeter with a measurement limit of 100 A and permissible absolute error of 1 A.

For measuring transducers, the errors can be represented relative to either the input or output signals. Let us consider the relationship between these two error representations.

Figure 2.1 shows the nominal and, let us assume, the real transfer functions of some transducer. The nominal transfer function, as done in practice whenever possible, is assumed to be linear. We denote the input quantity by x and the output quantity by y. They are related by the dependency

$$x = Ky,$$

where K is the nominal transduction constant.

At the point with true values of the quantities x_t and y_t, the true value of the transduction constant will be $K_t = x_t/y_t$. Calculations based on the nominal constant K, however, result in an error.

Let $x_a = Ky_t$ and $y_a = x_t/K$ be determined based on y_t and x_t (see Fig. 2.1).

Then the absolute transducer error with respect to the input will be

$$\Delta x = x_a - x_t = (K - K_1)y_t.$$

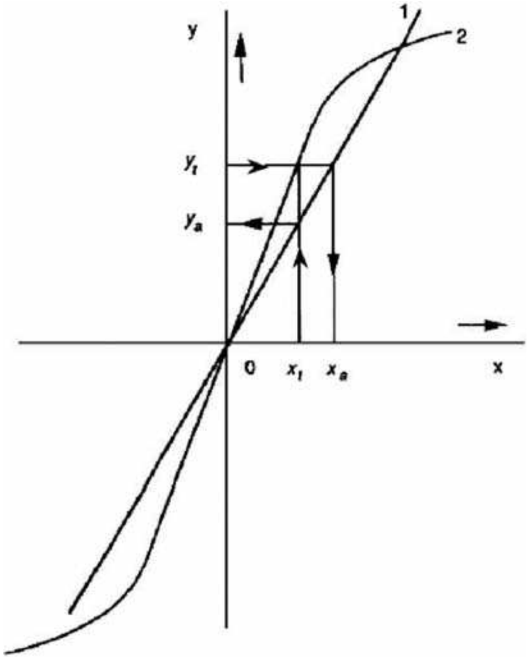


Fig. 2.1 Nominal (curve 1) and real (curve 2) transfer functions of a measuring transducer

The error with respect to the output is expressed analogously:

$$\Delta y = y_a - y_t = \left(\frac{1}{K} - \frac{1}{K_t}\right) x_t.$$

We note, first, that Δx and Δy always have different signs: If $(K - K_t) > 0$, then $(1/K - 1/K_t) < 0$.

But this is not the only difference. The quantities x and y can also have different dimensions; i.e., they can be physically different quantities, so that the absolute input and output errors are not comparable. For this reason, we shall study the relative errors:

$$\varepsilon_x = \frac{\Delta x}{x_t} = (K - K_t)\frac{y_t}{x_t} = \frac{K - K_t}{K_t},$$
$$\varepsilon_y = \frac{\Delta y}{y_t} = \frac{(K_t - K)}{KK_t}\frac{x_t}{y_t} = \frac{K_t - K}{K}.$$

As $K_t \neq K$, we have $|\varepsilon_x| \neq |\varepsilon_y|$.

We denote the relative error in the transduction constant at the point (x_t, y_t) as ε_k, where $\varepsilon_k = (K - K_t)/K_t$. Then

$$\frac{\varepsilon_x}{\varepsilon_y} = -(1 + \varepsilon_k).$$

However, $\varepsilon_k \ll 1$, and in practice relative errors with respect to the input and output can be regarded as equal in magnitude.

In measures, the rated error is determined as the difference between the nominal value of the measure and the "true value" of the quantity reproduced by the measure; the "true value" is obtained by another, known to be much more precise, measurement. This is analogous to indicating instruments if one considers the nominal value of a measure as the indication of the instrument.

It is interesting to note that single measures that reproduce passive quantities, for example, mass, electric resistance, and so on, have only systematic errors. The error of measures of active quantities (electric voltage, electric current, etc.) can have both systematic and random components. Multiple-valued measures of passive quantities can have random errors introduced by the switching elements.

To summarize, when the errors of measuring instruments are rated, the permissible limits of the intrinsic and all additional errors are specified. At the same time, the reference and rated operating conditions are indicated.

Of all forms enumerated above for expressing the errors of measuring instruments, the best is the relative error, because in this case, the indication of the permissible limit of error gives the clearest idea of the level of measurement accuracy that can be achieved with the given measuring instrument. The relative error, however, usually changes significantly over the measurement range of the instrument, and for this reason, it is difficult to be rated.

The absolute error is frequently more convenient than the relative error. In the case of an instrument with a scale, the limit of the permissible absolute error can be rated with the same numerical value for the entire scale of the instrument. But then it is difficult to compare the accuracies of instruments having different measurement ranges. This difficulty disappears when the fiducial errors are rated.

Let us now consider how the limits of permissible errors are expressed. For our discussion below, we shall follow primarily [9]. The limit of the permissible absolute error can sometimes be expressed by a single value (neglecting the sign):

$$\Delta = \pm a,$$

sometimes in the form of the linear dependence:

$$\Delta = \pm(a + bx), \tag{2.1}$$

where x is the nominal value of the measure, the indication of a measuring instrument, or the signal at the input of a measuring transducer, and a and b are constants, and sometimes by a general equation,

$$\Delta = f(x).$$

When the last dependence is complicated, it is given in the form of a table or graph.

The fiducial error γ (in percent) is defined by the formula

$$\gamma = 100\Delta/x_N,$$

where x_N is the fiducial value.

The fiducial value is assumed to be equal to the following:

1. The value at the end of the instrument scale.
2. The nominal value of the measurand, if it has been established.
3. The length of the scale, if the scale graduations narrow sharply toward the end of the scale. In this case, the error and the length of the scale are expressed in the same units of length (e.g., centimeters).
4. The rules above are in accordance with Recommendation 34 of OIML [9]. However, Publication 51 of IEC [8] foresees that if the zero marker falls within the scale, the fiducial value is equal to the (the span of the scale), which is a sum of the end values of the scale (neglecting their signs). This is controversial and we will discuss it in detail below.

A better between these two recommendations is the one by OIML. Indeed, consider, for example, an ammeter with a scale –100–0–100 A and with a permissible absolute error of 1 A. In this case, the fiducial error of the instrument will be 1% according to OIML and 0.5% according to IEC. But when using this instrument, the possibility of performing a measurement with an error of up to 0.5% cannot be guaranteed for any point of the scale, which makes the interpretation of the fiducial error confusing. An error not exceeding 1%, however, can be guaranteed when measuring a current of 100 A under reference conditions.

The tendency to choose a fiducial value such that the fiducial error would be close to the relative error of the instrument was observed in the process of improving IEC Publication 51. Indeed, in the previous edition of this publication, the fiducial value for instruments without a zero marker on the scale was taken to be equal to the difference of the end values of the range of the scale, and now it is taken to be equal to the larger of these values (neglecting the sign). Consider, for example, a frequency meter with a scale 45–50–55 Hz and the limit of permissible absolute error of 0.1 Hz. According to the previous edition of EIC Publication 51, the fiducial error of the frequency meter was assumed to be equal to 1%, and the current edition makes it equal to 0.2%. But when measuring the nominal 50-Hz frequency, the instrument relative error indeed will not exceed 0.2% (under reference conditions), while the 1% error has no relation to the accuracy of this instrument. Thus, the current edition is better. We hope that IEC will take the next step in this direction and take into account the recommendation of OIML for setting the fiducial value of instruments with a zero marker within the scale.

The limits of permissible relative error are rarely listed as rated but can be computed. If the rated error is expressed as the fiducial error γ (in percent), the permissible relative error for each value of the measurand must be calculated according to the formula

$$\delta = \gamma \frac{x_N}{x}.$$

If the rated error is expressed as the limits of absolute error Δ, the limit of permissible relative error δ is usually expressed in percent according to the formula

$$\delta = \frac{100\Delta}{x} = \pm c.$$

For digital instruments, the errors are often rated in the conventional form

$$\pm (b + q), \tag{2.2}$$

where b is the relative error in percent and q is some figure of the least significant digit of the digital readout device. For example, consider a digital millivoltmeter with a measurement range of 0–300 mV and with the indicator that has four digits. The value of one unit in the least significant digit of such an instrument is 0.1 mV. If this instrument is assigned the limits of permissible error $\pm(0.5\% + 2)$, then figure 2 in the parentheses corresponds to 0.2 mV. Now the limit of the relative error of the instrument when measuring, for example, a voltage of 300 mV can be calculated as follows:

$$\delta = \pm\left(0.5 + \frac{0.2 \times 100}{300}\right) = \pm 0.57\%.$$

Thus, to estimate the limit of permissible error of an instrument from the rated characteristics, some calculations must be performed. For this reason, although the conventional form (2.2) gives a clear representation of the components of instrument error, it is inconvenient to use.

A more convenient form is given in Recommendation 34 of OIML: According to this recommendation, the limit of permissible relative error is expressed by the formula

$$\delta = \pm\left[c + d\left(\frac{x_e}{x} - 1\right)\right], \tag{2.3}$$

where x_e is the end value of the measurement range of the instrument or the input signal of a transducer and c and d are relative quantities.

In (2.3), the first term on the right-hand side is the relative error of the instrument at $x = x_e$. The second term characterizes the increase of the relative error as the indications of the instrument decrease.

Equation (2.3) can be obtained from (2.2) as follows. To the figure q, there corresponds the measurand qD, where D is the value of one unit in the least significant digit of the instrument's readout device. In the relative form, it is equal to qD/x. Now, the physical meaning of the sum of the terms b and qD/x is that it is the limit of permissible relative error of the instrument. So,

$$\delta = \left(b + \frac{qD}{x}\right).$$

Using identity transformation, we obtain

$$\delta = b + \frac{qD}{x} + \frac{qD}{x_e} - \frac{qD}{x_e} = \left(b + \frac{qD}{x_e}\right) + \frac{qD}{x_e}\left(\frac{x_e}{x} - 1\right).$$

If we denote

$$c = b + \frac{qD}{x_e}, \quad d = \frac{qD}{x_e},$$

we obtain (2.3).

In application to the example of a digital millivoltmeter studied above, we have

$$\delta = \pm\left[0.57 + 0.07\left(\frac{x_e}{x} - 1\right)\right].$$

It is clear that the last expression is more convenient to use, and in general, it is more informative than the conventional expression (2.2).

Note that for standardization, the error limits are established for the total instrument error and not for the separate components. If, however, the instrument has an appreciable random component, then permissible limits for it are established separately, in addition to the limits of the total error. For example, aside from the limits of the permissible intrinsic error, the limits of the permissible dead band are also established.

Additional errors (recall that these are errors due to the deviation of the corresponding influence quantities from their values falling within the reference condition) of measuring instruments are rated by prescribing the limits for each additional error separately. The intervals of variation of the corresponding influence quantities are indicated simultaneously with the limits of the additional errors. The collection of ranges provided for all influence quantities determines the rated operating conditions of the measuring instrument. The limits of permissible additional errors are often represented in proportion to the values of their corresponding influence quantities or the deviation of these quantities from the limits of the intervals determining their reference values. In this case, the corresponding coefficients are rated. We call them the influence coefficients.

In the case of indicating measuring instruments, additional errors are often referred to by the term *variation of indications.* This term is used, in particular, for electric measuring instruments [8].

The additional errors arising when the influence quantities are fixed are systematic errors. For different instruments of the same type, however, systematic errors can have different values and, moreover, different signs. For this reason, the documentation for the overwhelming majority of instrument types sets the limits of additional errors as both positive and negative with equal numerical values. For example, the change in the indications of an electric measuring instrument caused by a change in the temperature of the surrounding medium should not exceed the limits $\pm 0.5\%$ for each 10°C change in temperature under rated operating conditions (the numbers here are arbitrary).

If, however, the properties of different measuring devices of a given type are sufficiently uniform, it is best to standardize the influence function, i.e., to indicate the dependence of the indications of the instruments or output signals of the transducers on the influence quantities and the limits of permissible deviations from each such dependence. If the influence function can be standardized, then it is possible to introduce corrections to the indications of the instruments and thereby to use the capabilities of the instruments more fully.

Figure 2.2 shows how the instrument errors depend on the values of an influence quantity, assuming two basic alternatives for rating the additional errors. The upper figure represents the case where the documentation lists the limits of the intrinsic and additional errors. Such rating stipulates that the instrument accuracy is determined by the limits of the intrinsic error as long as the influence quantity is within reference condition and by the sum of the limits of the intrinsic and constant limits of the additional errors if the influence quantity is within rated operating condition. The lower figure depicts the case when the documentation lists the limits of the intrinsic error and the influence coefficients for the additional errors. Here, when

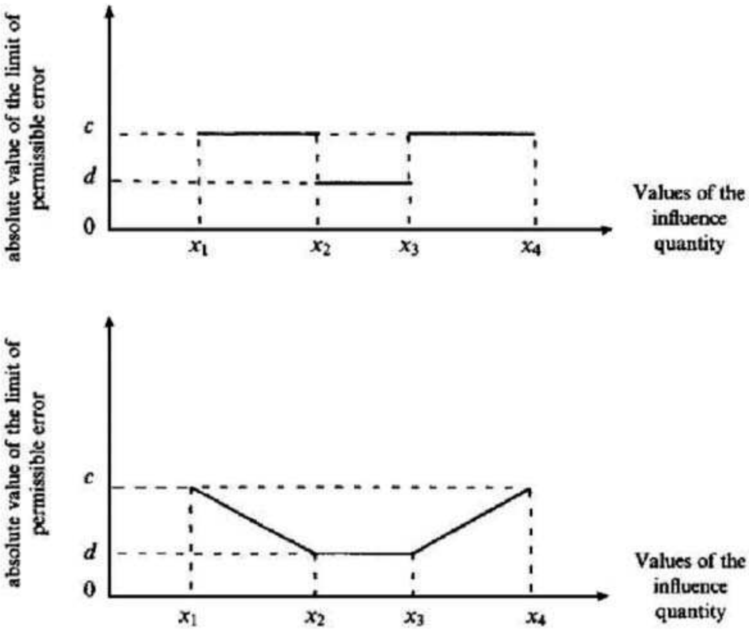


Fig. 2.2 Two variants of rating limits of additional errors of measuring instruments. The interval (x_2, x_3) corresponds to reference conditions; the interval (x_1, x_4) corresponds to the rated operating conditions; d is the absolute value of the limits of permissible intrinsic error; c is the absolute value of the limits of permissible error in the rated operating conditions; and $(c–d)$ is the absolute value of the limits of permissible additional error

the influence quantity is outside the reference condition, the limits of the additional error expand linearly with the deviation of the influence quantity from the reference condition (as long as the influence quantity stays within the rated operating conditions).

It should be emphasized that the actual additional errors that can arise in a measurement will depend not only on the properties of the measuring instrument but also on the accuracy of obtaining the values of the corresponding influence quantities.

Often a measuring instrument has an electrical signal on its input. This input signal can be characterized by several parameters. One of them reflects the magnitude of the measurand. This parameter is called the *informative parameter*: By measuring its magnitude, we can find the value of the measurand. All other parameters do not have direct connections with the magnitude of the measurand, and they are called *noninformative parameters.*

Measuring instruments are constructed with the goal to make them insensitive to all noninformative parameters of the input signal. This goal, however, cannot be achieved completely, and in the general case, the effect of the noninformative parameters can only be decreased but not eliminated. But, for all noninformative parameters, it is possible to determine limits such that when the noninformative parameters vary within these limits, the total error of the measuring instrument will change insignificantly, which makes it possible to establish the reference ranges of the noninformative parameters.

If some noninformative parameter falls outside the reference limits, then the error arising is regarded as another additional error. The effect of each noninformative parameter is rated separately, as for influence quantities. Furthermore, rating the additional errors arising from noninformative parameters is done based on the same assumptions as those used for rating the additional errors caused by the influence quantities.

The errors introduced by changes in the noninformative parameters of the input signals are occasionally called *dynamic errors.* In the presence of multiple parameters, however, this term is not expressive. It is more intuitive to give each error a characteristic name, as is usually done in electric and radio measurements. For example, the change in the indications of an AC voltmeter caused by changes in the frequency of the input signal is called the frequency error. In the case of the measurements of the peak variable voltages, apart from the frequency errors, the errors caused by changes in the widths of the pulse edges, the decay of the flat part of the pulse, and so on are called the shape errors.

Another property of measuring instruments that affects their accuracy and is also rated is stability. Stability, like accuracy, is a positive quality of a measuring instrument. Just as the accuracy is characterized by inaccuracy (error, uncertainty), stability is characterized by instability. An important particular case of instability is drift. Drift is usually not rated. Instead, when it is discovered, the zero indication of the instrument is reset.

The first method of rating the instability involves stipulating the time period after which the instrument must be checked and calibrated if needed. The second method consists of indicating different limits for the error of the instrument for different periods of time after the instrument was calibrated. For example, the following table (taken with modifications from [18]) can be provided in the specifications of a digital instrument:

Time after calibration	24 h	3 months	1 year	2 years
Temperature	$23 \pm 1°C$	$23 \pm 5°C$	$23 \pm 5°C$	$23 \pm 5°C$
Limits of error	$\pm(0.01\%$ $+1$ unit)	$\pm(0.015\%$ $+1$ unit)	$\pm(0.02\%$ $+1$ unit)	$\pm(0.03\%$ $+2$ units)

In the last line entries, the first number in the parentheses specifies the percent of the instrument indication and the second is a figure of the least significant digit (from 0 to 9). The second number lists the absolute error in units of the least significant digit of the instrument. To find the corresponded part of the limits of error of that instrument, one must calculate the value of this number in units of measurement. For example, if the above table is given in the documentation of a millivoltmeter with the range of 300 mV and 4-digit readout device, then the value of the least-significant digit is 0.1 mV. Assume that a user utilizes this instrument 2 years after calibration and the readout is 120.3 mV. Then, the limits of error of this instrument for this measurement are $\pm(120.3 \times 0.0003 + 0.2) = \pm 0.24\,\text{mV}$. The second number is constant for a given instrument range. It was called the *floor error* in [18].

Obviously, specifying how instrument accuracy changes with time since calibration conveys more information about the instrument characteristics than simply rating the interval between calibrations, and this extra information is beneficial to the users.

Below is another example of specification of a digital multirange voltmeter, also from [18] (the specification for only two ranges is shown).

Time after calibration	24 h	90 days	12 months	Temperature coefficient
Temperature	23 ± 1°C	23 ± 5°C	23 ± 5°C	0–18 & 28–55°C Per 1°C
10.00000 V	–	–	±(35ppm +5ppm)	±(5ppm +1ppm)
1000.000 V	±(20ppm +6ppm)	±(35ppm +10ppm)	±(45ppm +10ppm)	±(5ppm + 1ppm)

The last two rows in the above table give the limits of error of the instrument depending on the time from the calibration. The numbers in parentheses specify limits of two additive parts of the error in ppm. A confusing aspect here is that the first part is expressed as a relative error since the first number gives the limits of error relative to the indication of the instrument for a given measurement, while the second number specifies the error relative to the instrument range, the same as the floor error in the previous example.

The last column specifies the limits of the additional error due to temperature deviation from reference conditions. These limits are rated in the form shown in the lower graph of Fig. 2.2: the limits of the additional error grow by the specified amount for each 1°C of temperature deviation.

We provide examples of using this table in Sect. 4.6 for a measurement under reference temperature conditions and in Sect. 4.7 for a measurement under rated conditions.

The above excerpts of instrument specifications show the importance of understanding conventions used by the manufacturer of the instrument in specifying the instrument accuracy in its certificate. This is especially true if the manufacturer does not follow recommendations for rating the accuracy of instruments that have been issued by organizations such as OIML.

Rating of errors predetermines the properties of measuring instruments and is closely related with the concept of *accuracy classes* of measuring instruments. The purpose of this concept is the unification of the accuracy requirements of measuring instruments, the methods for determining them, and the accuracy-related notation in general, which is certainly useful to both the manufacturers of measuring instruments and to users. Indeed, such unification makes it possible to limit, without harming the manufacturers or the users, the list of instruments, and it makes it easier to use and check the instruments. We shall now discuss this concept in greater detail.

Accuracy classes were initially introduced for indicating electric measuring instruments [8]. Later this concept was also extended to all other types of measuring instruments [9]. In [1], the following definition is given for the term accuracy class: The accuracy class is a class of measuring instruments or measuring systems that meet certain stated metrological requirements intended to keep instrumental errors or uncertainties within specified limits under specified operating conditions.

Unfortunately, this definition does not entirely reflect the meaning of this term. Including measurement systems into the definition is incorrect because systems are usually unique and thus are not divided into classes. Further, instrumental errors and uncertainties are properties of measurements – not instruments – and hence should not be used to define instrument classes. A better definition is given in the previous edition of VIM: The accuracy class is a class of measuring instruments that meets certain metrological requirements that are intended to keep errors within specified limits.

Every accuracy class has conventional notation, established by agreement – the class index – that is presented in [8, 9]. On the whole, the accuracy class is a generalized characteristic that determines the limits for all errors and all other characteristics of measuring instruments that affect the accuracy of measurements performed with their help.

For measuring instruments whose permissible limits of intrinsic error are expressed in the form of relative or fiducial errors, the following series of numbers, which determine the limits of permissible intrinsic errors and are used for denoting the accuracy classes, was established in [9]:

$$(1, 1.5, 1.6, 2, 2.5, 3, 4, 5, \text{ and } 6) \times 10^n,$$

where $n = +1, 0, -1, -2, \ldots$; the numbers 1.6 and 3 can be used, but are not recommended. For any one value of n, not more than five numbers of this series (i.e., no more than five accuracy classes) are allowed. The limit of permissible intrinsic error for each type of measuring instrument is set equal to one number in the indicated series.

Table 2.1 gives examples of the adopted designations of accuracy classes of these measuring instruments.

In those cases when the limits of permissible errors are expressed in the form of absolute errors, the accuracy classes are designated by Latin capital letters or roman numerals. For example, [41] gives the accuracy classes of block gauges as Class X, Y, and Z. Gauges of Class X are the most accurate; gauges of Class Y are less accurate than Class X, and gauges of Class Z are the least accurate.

If (2.3) is used to determine the limit of permissible error, then both numbers c and d are introduced into the designation of the accuracy class. These numbers are selected from the series presented above, and in calculating the limits of permissible error for a specific value of x, the result is rounded so that it would be expressed by not more than two significant digits.

Table 2.1 Designations of accuracy classes

Form of the expression for the error	Limit of permissible error (examples)	Designation of the accuracy class (for the given example)
Fiducial error, if the fiducial value is expressed in units of the measurand	$\gamma = \pm 1.5\%$	1.5
Fiducial error, if the fiducial value set equal to the scale length	$\gamma = \pm 0.5\%$	0.5
Relative error, constant	$\delta = \pm 0.5\%$	0.5
Relative error, increasing as the measurand decreases	$\delta = \pm \left[0.02 + 0.01\left(\frac{x_e}{x} - 1\right)\right]\%$	0.02/0.01

In conclusion, we shall formulate the basic rules for rating errors of measuring instruments:

1. All properties of a measuring instrument that affect the accuracy of the results of measurements must be rated.
2. Every property that is to be rated should be rated separately.
3. Rating methods must make it possible to check experimentally, and as simply as possible, how well each individual measuring instrument corresponds to the established requirements.

In some cases, exceptions must be made to these rules. In particular, an exception is necessary for strip strain gauges that can be glued on an object only once. Since these strain gauges can be applied only once, the gauges that are checked can no longer be used for measurements, whereas those that are used for measurements cannot be checked or calibrated.

In this case, it is necessary to resort to regulation of the properties of a *collection* of strain gauges, such as, for example, the standard deviation of the sensitivity and mathematical expectation of the sensitivity. The sensitivity of a particular strain gauge, which is essentially not a random quantity in the separate device, is a random quantity in a collection of strain gauges. Since we cannot check all the gauges, a random sample, representing a prescribed p percent of the entire collection (which could be, e.g., all gauges produced in a given year), is checked. Once the sensitivity x_i of every selected gauge has been determined, it is possible to construct a *statistical tolerance interval*, i.e., the interval into which the sensitivity of any random sample of p percent of the entire collection of strain gauges will fall with a chosen probability α. As $\alpha \neq 1$ and $p \neq 1$, there is a probability that the sensitivity of any given strain gauge falls outside these tolerance limits. For this reason, the user must take special measures that address such a case. In particular, several strain gauges, rather than one, should be used.

2.4 Dynamic Characteristics of Measuring Instruments

The dynamic characteristics of measuring instruments reflect the relation between the change in the output signal and an action that produces this change. The most important such action is a change in the input signal. In this case, the dynamic characteristic is called the dynamic characteristic for the input signal. Dynamic characteristics for various influence quantities and for a load (for measuring instruments whose output signal is an electric current or voltage) are also studied.

Complete and partial dynamic characteristics are distinguished [27].

The complete dynamic characteristics determine uniquely the change in time of the output signal caused by a change in the input signal or by other action. Examples of such characteristics include a differential equation, transfer function, amplitude- and phase-frequency response, and the transient response. These characteristics are essentially equivalent, but the differential equation is the basic characteristic from which the other characteristics are derived.

A partial dynamic characteristic is a parameter of the full dynamic characteristic (introduced shortly) or the response time of the instrument. Examples are the response time of the indications of an instrument and the transmission band of a measuring amplifier.

Measuring instruments[1] can most often be regarded as inertial systems of first or second order. If $x(t)$ is the signal at the input of a measuring instrument and $y(t)$ is the corresponding signal at the output, then the relation between them can be expressed with the help of first-order (2.4) or second-order (2.5) differential equations, respectively, which reflect the dynamic properties of the measuring instrument:

$$Ty'(t) + y(t) = Kx(t), \tag{2.4}$$

$$\frac{1}{\omega_0^2} y''(t) + \frac{2\beta}{\omega_0} y'(t) + y(t) = Kx(t). \tag{2.5}$$

The parameters of these equations have specific names: T is the time constant of a first-order device, K is the transduction coefficient in the static state, ω_0 is the angular frequency of free oscillations, and β is the damping ratio. An example of a real instrument whose properties are specified by the second-order differential equation is a moving-coil galvanometer. In this instrument type, $\omega_o = 2\pi\, 1/T_0$, where T_0 is the period of free oscillations (the reverse of the natural frequency) and β is the damping ratio, which determines how rapidly the oscillations of the moving part of the galvanometer will subside.

Equations (2.4) and (2.5) reflect the properties of real devices, and for this reason, they have zero initial conditions: for $t \le 0$, $x(t) = 0$ and $y(t) = 0$, $y'(t) = 0$ and $y''(t) = 0$.

[1] The rest of this section requires familiarity with control theory. The reader can skip this portion without affecting the understanding of the rest of the book.

To obtain transfer functions from differential equations, it is first necessary to move from signals in the time domain to their Laplace transforms, and then to obtain the ratio of the transforms. Thus,

$$\mathcal{L}[x(t)] = x(s), \qquad \mathcal{L}[y(t)] = y(s),$$
$$\mathcal{L}[y'(t)] = sy(s), \qquad \mathcal{L}[y''(t)] = s^2 y(s),$$

where s is the Laplace operator.

For the first-order system, in accordance to (2.4), we obtain

$$W(s) = \frac{y(s)}{x(s)} = \frac{K}{1 + sT},$$

and for the second-order system, from (2.5), we obtain

$$W(s) = \frac{y(s)}{x(s)} = \frac{K}{\left(1/\omega_0^2\right) s^2 + (2\beta/\omega_0)s + 1}. \tag{2.6}$$

Let us consider the second-order equation in more detail. If in the transfer function the operator s is replaced by the complex frequency $j\omega (s = j\omega)$, then we obtain the complex frequency response. We shall now study the relation between the named characteristics for the example of a second-order system. From (2.5) and (2.6), we obtain

$$W(j\omega) = \frac{K}{\left(1 - \omega^2/\omega_0^2\right) + j2\beta\omega/\omega_0}, \tag{2.7}$$

where $\omega = 2\pi f$ is the running angular frequency.

The complex frequency response is often represented with its real and imaginary parts,

$$W(j\omega) = P(\omega) + j\,Q(w).$$

In our case,

$$P(\omega) = \frac{K(1 - (\omega^2/\omega_0^2))}{(1 - (\omega^2/\omega_0^2))^2 + 4\beta^2(\omega^2/\omega_0^2)},$$

$$Q(\omega) = \frac{2\beta(\omega/\omega_0)K}{(1 - (\omega^2/\omega_0^2))^2 + 4\beta^2(\omega^2/\omega_0^2}.$$

The complex frequency response can also be represented in the form

$$W(j\omega) = A(\omega)e^{j\varphi(\omega)},$$

where $A(\omega)$ is the amplitude-frequency response and $\varphi(\omega)$ is the frequency response of phase. In the case at hand,

$$A(\omega) = \sqrt{P^2(\omega) + Q^2(\omega)} = \frac{K}{\sqrt{(1 - (\omega^2/\omega_0^2))^2 + 4\beta^2(\omega^2/\omega_0^2)}},$$
$$\varphi(\omega) = \arctan\frac{Q(\omega)}{P(\omega)} = -\arctan\frac{2\beta(\omega/\omega_0)}{1 - (\omega^2/\omega_0^2)}. \quad (2.8)$$

Equation (2.8) has a well-known graphical interpretation using the notion of transient response. The transient response is the function $h(t)$ representing the output signal produced by a unit step function $1(t)$ at the input. (The unit step function, which we denote $1(t)$, is a function whose value is 0 for $t < 0$ and 1 for $t \geq 0$.) As the input is not periodic, $h(t)$ is calculated with (2.4) or (2.5). Omitting the technical but, unfortunately, complicated calculations, we arrive at the final form of the transient response of the instrument under study:

$$h(t) = \begin{cases} 1 - e^{-\beta\tau}\dfrac{1}{\sqrt{1-\beta^2}}\sin\left(\tau\sqrt{1-\beta^2} + \arctan\dfrac{\sqrt{1-\beta^2}}{\beta}\right) & \text{if } \beta < 1, \\ 1 - e^{-\tau}(\tau + 1) & \text{if } \beta = 1, \\ 1 - e^{-\beta\tau}a\dfrac{1}{\sqrt{\beta^2-1}}\sinh\left(\tau\sqrt{\beta^2-1} + \operatorname{arctanh}\dfrac{\sqrt{\beta^2-1}}{\beta}\right) & \text{if } \beta > 1. \end{cases}$$

(Note that the cast case utilizes hyperbolic trigonometric functions.) In this expression, $\tau = \omega_0 t$ is normalized time, and the output signal is normalized to make its steady-state value equal to unity, i.e., $h(t) = y(t)/K$. Thus, the formulas above and the corresponding graphs presented in Fig. 2.3 are universal in the sense that they do not depend on the specific values of ω_0 and K.

It should be noted that some types of measuring instruments do not have dynamic characteristics at all; these include measures of length, weights, vernier calipers, and so on. Some measuring instruments, such as measuring capacitors (measures of capacitance), do not have an independent dynamic characteristic by themselves. But when they are connected into an electric circuit, which always has some resistance and sometimes an inductance, the circuit always acquires, together with a capacitance, definite dynamic properties.

Measuring instruments are diverse. Occasionally, to describe adequately their dynamic properties, it is necessary to resort to linear equations of a higher order,

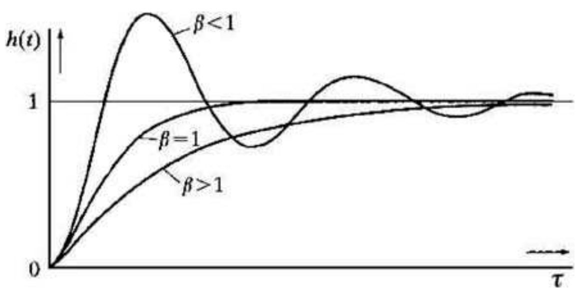


Fig. 2.3 The transient response of an instrument described by a second-order differential equation; β is the damping ratio

nonlinear equations, or equations with distributed parameters. However, complicated equations are used rarely, and it is not an accident. After all, measuring instruments are created specially to perform measurements, and their dynamic properties are made to guarantee convenience of use. For example, in designing a recording instrument, the transient response is made to be short, approaching the steady state level monotonically or oscillating insignificantly. In addition, the scale of the recording instrument is typically made to be linear. But when these requirements are met, the dynamic properties of the instrument can be described by one characteristic corresponding to a linear differential equation of order no higher than second.

Rating of the dynamic characteristics of measuring instruments is performed in two stages. First, an appropriate dynamic characteristic to be rated must be chosen, and second, the nominal dynamic characteristic and the permissible deviations from it must be established.

For recording instruments and universal measuring transducers, a complete dynamic characteristic, such as transient response, must be rated: Without having the complete dynamic characteristic, a user cannot effectively use these instruments.

For indicating instruments, it is sufficient to rate the response time. In contrast to the complete characteristics, this characteristic is a partial dynamic characteristic. The dynamic error is another form of a partial dynamic characteristic. Rating the limits of a permissible dynamic error is convenient for the measuring instruments employed, but it is justified only when the shape of the input signals does not change much.

For measuring instruments described by linear first- and second-order differential equations, the coefficients of all terms in the equations can be rated. In the simplest cases, the time constant is rated in the case of a first-order differential equation, and the natural frequency and the damping ratio of the oscillations are standardized in the case of a second-order differential equation.

When imposing requirements on the properties of measuring instruments, it is always necessary to keep in mind how compliance will be checked. For dynamic characteristics, the basic difficulties have to do with creating test signals of predetermined form (with sufficient accuracy), or with recording the input signal with a dynamically more accurate measuring instrument than the measuring instrument whose dynamic properties are being checked.

If adequately accurate test signals can be created and used to obtain the dynamic characteristic, i.e., a transient response as a response of a unit step function signal and frequency response as a response of a sinusoidal test signal, then in principle the instrument can be checked without any difficulties.

But sometimes the problem must be solved with a test signal that does not correspond to the signal intended for determining the complete dynamic characteristic. For example, one would think that tracing of signals at the input and output of a measuring instrument could solve the problem. In this case, however, special difficulties arise: small errors in recording the test signal and reading the values of the input and output signals often render the dynamic characteristic obtained from them physically meaningless and not corresponding to the dynamic properties of the measuring instrument. Such an unexpected effect occurs because the problem at hand is a so-called improperly posed problem. A great deal of attention is currently being devoted to such problems in mathematics, automatics, geophysics, and other disciplines. Improperly posed problems are solved by the methods of regularization, which essentially consist of the fact that the necessary degree of filtering (smoothing) of the obtained solution is determined based on a priori information about the true solution. Improperly posed problems in dynamics in application to measurement engineering are reviewed in [27].

A separate problem, which is important for some fields of measurement, is the determination of the dynamic properties of measuring instruments directly when the instruments are being used. An especially important question here is the question of the effect of random noise on the accuracy with which the dynamic characteristics are determined.

This section, then, has been a brief review of the basic aspects of the problem of rating and determining the dynamic properties of measuring instruments.

2.5 Calibration and Verification of Measuring Instruments

Every country wishes to have trustworthy measurements. One of the most important arrangements to achieve this goal is to have a system for keeping errors of all measuring instruments within permissible limits. Therefore, all measuring instruments in use are periodically checked. In the process, working standards are used either to verify that the errors of the measuring instruments being checked do not exceed their limits or to recalibrate the measuring instruments.

The general term for the above procedures is *calibration.* But one should distinguish between a real calibration and a simplified calibration.

Real calibration results in the determination of a relation between the indications of a measuring instrument and the corresponding values of a working measurement standard. This relation can be expressed in the form of a table, a graph, or a function. It can also be expressed in the form of the table of corrections to the indications of the measuring instrument. In any case, as the result of real calibration, the indications of the working standard are mapped to the instrument being calibrated. Consequently, the accuracy of the instrument becomes close to the accuracy of the working standard.

Real calibration can be expensive, complex, and time-consuming.

Therefore, calibration is mostly used for precise and complex instruments. For other instruments, the simplified calibration suffices.

The simplified calibration (also called *verification*) simply reveals whether the errors of a measuring instrument exceed their specified limits. Essentially, verification is a specific case of quality control, much like quality control in manufacturing. And because it is quality control, verification results do have some rejects.

Further, verification can take the form of a complete or element-wise check. A complete check determines the error of the measuring instrument as a whole.

In the case of an element-wise check, the errors of the individual elements comprising the measuring instrument are determined. The overall error of the measuring instrument is then calculated using methods that were examined in [44].

A complete check is always preferable as it gives the most reliable result. In some cases, however, a complete check is impossible to perform and one must resort to an element-wise check. For example, element-wise calibration is often employed to check measuring systems when the entire system cannot be delivered to a calibration laboratory and the laboratory does not have necessary working standards that could be transported to the system's site.

The units of a system are verified by standard methods. When the system is verified, however, in addition to checking the units, it is also necessary to check the serviceability of the system as a whole. The methods for solving this problem depend on the arrangement of the system, and it is hardly possible to make general recommendations here. As an example, the following procedure can be used for a system with a temperature-measuring channel comprising a platinum–rhodium–platinum thermocouple as the primary measuring transducer and a voltmeter.

After all units of the system have been checked, we note the indication of the instrument at the output of the system. Assume that the indication is $+470°C$. For the most common types of thermocouples, there exists known standardized transfer function, while specific brands of thermocouple products have rated limits of deviation from the standardized function.

From the standardized transfer function of the primary measuring transducer, we obtain the output signal that should be observed for the given value of the measured quantity. For our thermocouple, when the temperature of $+470°C$ is measured, the EMF at the output of the thermocouple must be equal to 3.916 mV. Next, disconnecting the wires from the thermocouple and connecting them to the voltage exactly equal to the nominal output signal of the thermocouple, we once again note the indication of the voltmeter. If it remains the same or has changed within the limits of permissible error of the thermocouple and voltmeter, then the system is serviceable.

Of course, this method of checking will miss the case in which the errors of both the thermocouple and voltmeter are greater than their respective permissible errors but these errors mutually cancel. However, this result can happen only rarely. Moreover, such a combination of errors is in reality permissible for the system.

Let us now consider complete check verification in more detail. Here, the values represented by working standards are taken as true values, and the instrument indication is compared to these values. In fact, a working standard has errors. Therefore, some fraction of serviceable instruments, i.e., instruments whose errors do not exceed the limits established for them, is rejected in a verification – false rejection – and some fraction of instruments that are in reality unserviceable are accepted – false retention. This situation is typical for monitoring production quality, and just as with quality control, a probabilistic analysis of the procedure is useful to understand the extent of a potential issue.

Without loss of generality, suppose for simplicity that the complete check verification is performed by measuring the same quantity simultaneously using a working standard (which in this case is an accurate measuring instrument) and the instrument being checked. Accordingly, we have

$$A = x - \zeta = y - \gamma,$$

where A is the true value of the quantity, x and y are the indications of the instrument and working standard, and ζ and γ are the errors of the instrument and working standard. It follows from the above equation that the difference z between the indications of the instrument and the standard is equal to the difference between their errors,

$$z = x - y = \zeta - \gamma. \tag{2.9}$$

We are required to show that $|\zeta| \leq \Delta$, where Δ is the limit of permissible error of the instrument. From the experimental data (i.e., from the indications), we can find z; because γ is supposed to be much smaller than ζ, we shall assume that if $|z| \leq \Delta$, then the checked instrument is serviceable, and if $|z| > \Delta$, then it is not serviceable.

To perform probabilistic analysis of when the above assumption is wrong, it is necessary to know the probability distribution for the errors of the checked and standard instruments. Let us suppose we know these distributions. The probability of a false rejection is

$$p_1 = P\{|\zeta - \gamma| > \Delta|_{|\zeta|\leq\Delta}\},$$

and the probability of a false retention is

$$p_2 = P\{|\zeta - \gamma| \leq \Delta|_{|\zeta|>\Delta}\}.$$

A false rejection is obtained for $|\zeta| \leq \Delta$ when $|\zeta - \gamma| > \Delta$, i.e.,

$$\zeta - \gamma > \Delta, \zeta - \gamma < -\Delta,$$

or

$$\gamma < \zeta - \Delta, \gamma > \zeta + \Delta.$$

If the probability density functions of the errors of the instrument and working standard are $f(\zeta)$ and $\varphi(\gamma)$, respectively, then

$$p_1 = \int_{-\Delta}^{\Delta} f(\zeta)\left(\int_{-\infty}^{\zeta-\Delta} \varphi(\gamma)\,d\gamma + \int_{\zeta+\Delta}^{+\infty} \varphi(\gamma)\,d\gamma\right) d\zeta.$$

A false retention is possible when $|\zeta| > \Delta$, i.e., when $\zeta > +\Delta$ and $\zeta < -\Delta$.

In this case, $|\zeta - \gamma| \leq \Delta$, i.e.,

$$\zeta - \gamma \leq \Delta, \zeta - \gamma \geq -\Delta,$$

or

$$\zeta - \Delta \leq \gamma \leq \zeta + \Delta.$$

Therefore,

$$p_2 = \int_{-\infty}^{-\Delta} f(\zeta) \left(\int_{\zeta-\Delta}^{\zeta+\Delta} \varphi(\gamma)\, d\gamma \right) d\zeta + \int_{\Delta}^{+\infty} f(\zeta) \left(\int_{\zeta-\Delta}^{\zeta+\Delta} \varphi(\gamma)\, d\gamma \right) d\zeta.$$

Thus, if the probability densities are known, then the corresponding values of p_1 and p_2 can be calculated; one can furthermore understand how these probabilities depend on the difference between the limits of permissible errors of the instrument being checked and the working standard.

If, in addition, cost considerations are added, then one would think about the problem of choosing the accuracy of the working standard that would be suitable for checking a given instrument. In reality, when the accuracy of working standards is increased, the cost of verification increases also. A rejection also has a certain cost. Therefore, by varying the limits of error of working standards, it is possible to find the minimum losses, and this accuracy is regarded as optimal.

Mathematical derivations aside, it is unfortunately difficult to estimate the losses from the use of instruments whose errors exceed the established limits, when these instruments pass the verification. In general, it is hard to express in monetary terms the often-significant economic effect of increasing measurement accuracy. For this reason, it is only in exceptional cases that economic criteria can be used to justify the choice of the relation between the limits of permissible error of the working standard and the checked instruments.

In addition, as has already been pointed out above, the fundamental problem is to determine the probability distribution of the errors of the instruments and standards. The results, presented in Sect. 2.7 below, of the statistical analysis of data from the verification of a series of instruments showed that the sampling data of the instrument errors are statistically unstable. Therefore, the distribution function of the instrument errors cannot be found from these data. However, there are no other data; it simply cannot be obtained anywhere.

Thus, it is impossible to find a sufficiently convincing method for *choosing* the relation between the permissible errors of the working standard and the instrument to be checked. For this reason, in practice, this problem is solved by a volitional method, by *standardizing* the relation between the limits of permissible errors. In practice, the calibration laboratories accept that the accuracy of a working standard must be four times higher than the accuracy of the checked instrument [18,26]. This means that some instruments that pass the verification may have errors exceeding by 25% the permissible level. Yet more aggressive ratios between the limits of permissible errors of the standard and the instrument, such as 1:10, are usually technically difficult to achieve.

It turns out, however, that a change in the verification process can eliminate this problem. Let us consider this method.

By definition, a serviceable instrument is an instrument for which $|x - A| \leq \Delta$ and an instrument is unserviceable if $|x - A| > \Delta$. Analogous inequalities are also valid for a working standard: $|y - A| \leq \Delta_s$, if the instrument is serviceable and $|y - A| > \Delta_s$ if it is not serviceable.

For $x > A$, for a serviceable instrument, $x - A \leq \Delta$. But $y - \Delta_s \leq A \leq y + \Delta_s$. For this reason, replacing A by $y - \Delta_s$, we obtain for a serviceable instrument,

$$x - y \leq \Delta - \Delta_s. \tag{2.10}$$

Analogously, for $x < A$, for a serviceable instrument,

$$x - y \geq -(\Delta - \Delta_s). \tag{2.11}$$

Repeating the calculations for an unserviceable instrument, it is not difficult to derive the corresponding inequalities:

$$x - y > \Delta + \Delta_s, \tag{2.12}$$

$$x - y < -(\Delta + \Delta_s). \tag{2.13}$$

Figure 2.4 graphically depicts the foregoing relations. Let the scale of the checked instrument be the abscissa axis. On the ordinate axis, we mark the points $+\Delta$ and $-\Delta$, and around each of these points, we mark the points displaced from them by $+\Delta_s$ and $-\Delta_s$. If Δ and Δ_s remain the same for the entire scale of the instrument, then we draw from the marked points on the ordinate axis straight lines parallel to the abscissa axis.

Region I corresponds to inequalities (2.10) and (2.11). The instruments for which the differences $x - y$ fall within this region are definitely serviceable irrespective of the ratio of the errors of the standard and checked instruments. Inequalities (2.12) and (2.13) correspond to regions II and III. The instruments for which the differences $x - y$ fall within the regions II or III are definitely unserviceable.

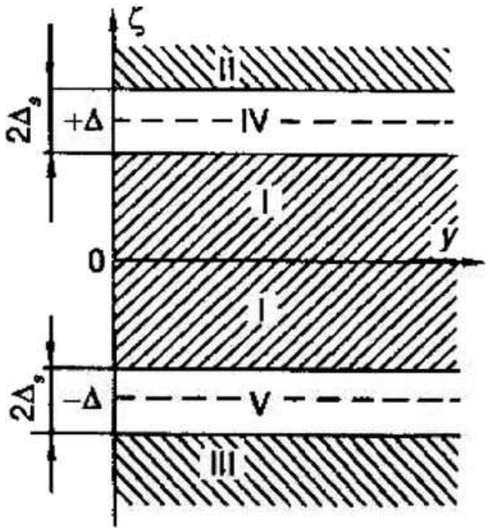


Fig. 2.4 Zones of definite serviceability (I), definite rejection (II and III), and uncertainty (IV and V) during verification of measuring instruments with the limit of permissible error Δ based on a working standard whose limit of permissible error is Δ_s

Some instruments can have errors such that

$$\Delta - \Delta_s < |x - y| < \Delta + \Delta_s.$$

These errors correspond to regions IV and V in Fig. 2.4. Such instruments essentially cannot be either rejected or judged to be serviceable, because in reality, they include both serviceable and unserviceable instruments. If they are assumed to pass verification, then the user will get some unserviceable instruments. This can harm the user. If, however, all such doubtful instruments are rejected, then in reality, some serviceable instruments will be rejected.

For instruments that are doubtful when they are manufactured or when they are checked after servicing, it is best that they be judged unserviceable. This tactic is helpful for the users and forces the manufacturers to employ more accurate standards to minimize the rejects. But this approach is not always practical. When the percentage of doubtful instruments is significant and the instruments are expensive and difficult to fix, it is best to check them again. Here, several variants are possible. One variant is to recheck the doubtful instruments with the help of more accurate working standards. When this is impossible, the verification can also be performed using other samples of working standards that are rated at the same accuracy as those used in the initial check. As different working standards have somewhat different errors, the results of comparing the instruments with them will be somewhat different. Thus, some doubtful instruments will move to the regions in Fig. 2.4 that allow definitive verification outcomes.

Ideally, the best way to deal with the doubtful instruments is to increase the accuracy of the working standard. However, the question then arises as to how much the accuracy of the standard instruments should be increased. If there are no technical limitations, then the accuracy of the working standard can be increased until the instrument can be judged as being either serviceable or unserviceable. However, the limits of permissible error of the standard instrument rarely need to be decreased beyond about 10 times less than the limit of permissible error of the instrument: The errors of instruments are usually not stable enough to be estimated with such high accuracy.

Rejection of instruments under verification is eliminated completely if instead of verification the instruments are recalibrated. The accuracy of the newly calibrated instrument can be almost equal to the accuracy of the working standard, which makes this method extremely attractive. The drawback of this method is that the result of a calibration is most often presented in the form of a table of corrections to the indications of the instrument, which is inconvenient for using the instrument.

2.6 Designing a Calibration Scheme

Calibration is a metrological operation whose goal is to transfer decreed units of quantities from a primary measurement standard to a measuring instrument. To protect the primary standards and to support calibration of large numbers of instruments, this transfer is performed indirectly, with the help of intermediate standards. In fact, intermediate standards may themselves be calibrated against primary standards not directly but through other intermediary standards. Thus, the sizes of units reproduced by primary standards are transferred to intermediary standards and through them to measuring instruments.

The hierarchical relations of standards with each other and with measuring instruments that are formed to support calibration can be represented as a *calibration scheme*. Note that the discussion in this section also fully applies to verification and *verification schemes*, which are the analog of calibration schemes in the context of verification. The standards at the bottom of the calibration schemes, which are used to calibrate measuring instruments, are called working standards; the intermediate standards, situated between the primary and working standards in the scheme, are called secondary standards. For the purpose of the discussion in this section, we will refer to secondary standards, working standards, and measuring instruments together as *devices*.

Measurement standards belonging to a calibration scheme are divided into ranks. The rank of a standard indicates the number of steps included in transferring the size of a unit from the primary measurement standard to a given standard, i.e., the number of standards on the path from this standard to the primary standard in the calibration scheme.

One of the most difficult questions arising in the construction of calibration schemes is the question of how many ranks of standards should be provided. Three main factors play a role in deciding this question: accuracy, cost, and capacity. As the number of ranks increases, the error with which the size of a unit is transferred to the measuring instrument increases, because some accuracy is lost at every calibration step. For this reason, to obtain high accuracy, the number of ranks of standards should be reduced to a minimum. On the other hand, the more the number of ranks the greater the overall capacity of the scheme in terms of the number of measuring instruments it can calibrate. In addition, the higher the accuracy of standards, the more expensive they are, and the more expensive they are to use. Thus, from the cost perspective, it is desirable to reduce the number of high-accuracy standards by increasing the number of ranks in the scheme.

One would think that it should be possible to find an economically optimal number of ranks of the calibration scheme. Such optimization, however, would require information about the dependence between the cost of the equipment and labor and the accuracy of calibration. This information is usually not available. For this reason, in practice, the optimal calibration schemes cannot be determined, and calibration schemes are commonly constructed in an ad hoc manner. However, a method proposed below allows designing a calibration scheme in a methodical way at least to satisfy its capacity requirements with the minimum number of ranks, and hence with the highest possible calibration accuracy. Accuracy constrains permitting; one can always then increase the number of ranks in the resulting scheme to reflect specific economic considerations.

Figure 2.5 shows a typical structure of a calibration scheme. In the simplest case, when all measuring instruments in the calibration scheme have similar accuracy,

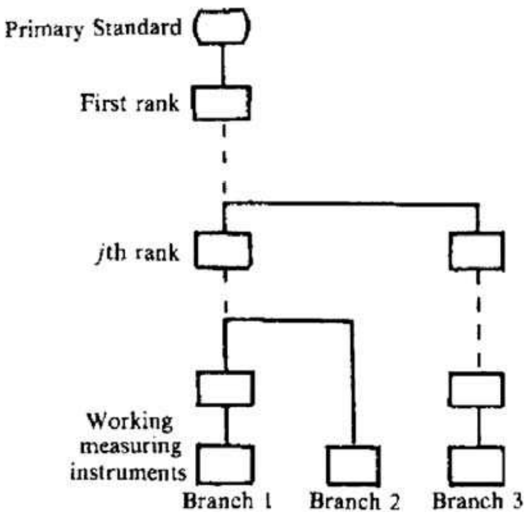


Fig. 2.5 A typical calibration scheme structure

a calibration scheme can be represented as a chain; for example, the entire calibration scheme on Fig. 2.5 would consist of just branch 1. The chain has the primary standard at the root, then certain number of secondary standards of the rank 1 below that are periodically calibrated against the primary standard, followed by a larger number of secondary standards of rank 2, each periodically calibrated against one of the standards of rank1, and so on until the measuring instruments at the leafs of the hierarchy.

However, some measuring instruments may be more accurate than others and cannot be calibrated by working standards at the bottom of the chain. These instruments must be "grafted" to the middle of the first branch, at the point where they can be calibrated by a standard of sufficient accuracy. These instruments form branch 2 on Fig. 2.5. The standard at the branching point in the calibration scheme acts as a secondary standard for one branch and a working standard for another.

Finally, there may be instruments of significantly different type than those in other branches, whose calibration requires some auxiliary devices between them and their working standards (such as scaling transducers in front of high-accuracy voltmeter for high voltage). The auxiliary devices introduce accuracy loss in calibration, and therefore they require the working standard to have a higher accuracy to account for this loss. In other words, if normally the accuracy ratio of the measuring instrument to working standard must be at most 1:4, (see Sect. 2.5 for the discussion on this accuracy relationship), this ratio must be lower (e.g., 1:10) for these instruments. To avoid the confusion, we place these instruments, along with the auxiliary devices, into distinct branches in the calibration scheme (such as branch 3 in Fig. 2.5). Such a branch can be grafted to another branch at an intermediary standard such that the ratio of its accuracy to the accuracy of the instruments corresponds to the requirement specific to the instruments' branch.

Secondary standards are usually calibrated with the highest possible accuracy, so that they can be also used as working standards for more accurate types of measuring instruments if needed. However, there is inevitable loss of accuracy with each calibration step. Consequently, different types of secondary standards are typically used for different ranks, and calibration at different ranks has different performance characteristics, such as time required to calibrate one device or time to prepare a standard for calibration (see below). At the same time, the types of devices that can be used at a given rank are usually known in advance, and it is only necessary to decide how many of them to procure and how to arrange them in an appropriate calibration scheme. Therefore, one can assume that the calibration frequency of secondary and working standards of a given rank, and how long each calibration takes, is known. Furthermore, we assume that the calibration frequency and time required to calibrate are known for all measuring instruments. Finally, the keepers of primary standards typically impose their own usage limits (e.g., they limit the number of calibrations that can be performed against the primary standard in one year). We assume that these limits are known as well.

We begin by considering the branch leading to the least accurate instruments as if it were the only branch in the scheme (e.g., branch 1 in Fig. 2.5). We call this branch a *stem.*

In such a single-branch calibration scheme, if the jth rank has N_j standards, then the maximum number of devices in the rank $(j+1)$ that can be supported will be

$$N_{j+1} = N_j \frac{\eta_j T_{j+1}}{t_{j+1}}, \tag{2.14}$$

where T_{j+1} is the time interval between calibrations of a device of rank $j+1$, t_{j+1} is the time necessary to calibrate one device in the rank $(j+1)$, and η_j is the utilization factor of the standards of rank j, considered below. Note that at the first calibration step, the number of secondary standards of rank 1 is determined as the minimum between the number given by (2.14) and the restrictions imposed by the keepers of the primary standards as mentioned earlier.

The utilization factor η_j reflects the fraction of time a corresponding standard can be used for calibration. In particular, η_j reflects the fact that the standard may only be used during the work hours; any losses of work time must also be taken into account. For example, if some apparatus is used 8 h per day and 1 h is required for preparation and termination, and preventative maintenance, servicing, etc. reduce the effective working time by 10%, then

$$\eta = \frac{8-1}{24} \times 0.9 = 0.2625.$$

Applying (2.14) to every step of the chain, we determine the capacity of the stem, which is the maximum number of standards of each rank and ultimately the number of measuring instruments $N_m^{(\max)}$ that can be supported by this calibration chain:

$$N_m^{(\max)} = N_0^{(\max)} N_1^{(\max)} \cdots N_{m-1}^{(\max)} = \prod_{j=0}^{m-1} \eta_j \frac{T_{j+1}}{t_{j+1}}, \tag{2.15}$$

where m is the total number of steps in transferring the size of a unit from the primary standard to the measuring instrument, inclusively and $N_j^{(\max)}$ is the maximum number of devices at each rank that a "full" calibration scheme can have.

On the other hand, to design a calibration chain, that is, to decide on the number of ranks in the calibration chain that can support a given number N_{instr} of instruments, one can use the following procedure.

To protect the primary standards, they are never used to calibrate the working standards directly. Thus, at least one rank of secondary standards is always needed. We compute the maximum number of the secondary standards of rank 1 N_1, which could be calibrated against the primary standard in our calibration chain, using (2.14). Next, we check using (2.14) again, if N_1 secondary standards can support calibration of N_{instr} instruments. If not, we know that we need more ranks in the calibration scheme.

In the latter case, we first check if the accuracy of the secondary standards of the new rank will still be sufficient to calibrate the instruments, given the instruments' accuracy. If not, we have to assume that the calibration of the given number of instruments is impossible with the required calibration frequency (this outcome is extremely rare in practice). Otherwise, we apply (2.14) again to compute the maximum number of secondary standards of rank 2, N_2, which can be supported by N_1 standards of rank 1. [Note that we apply (2.14) twice because the calibration time of a measuring instrument and secondary standard can be – and typically is – different]. We continue in this manner until we find the smallest number of ranks of secondary standards that can support N_{instr} measuring instruments.

We should mention that, after each iteration of the above algorithm, if the resulting capacity of the calibration scheme is close to required, an alternative to increasing the number of ranks is to raise the efficiency of calibration. This could be achieved by either increasing standard utilization η_j or by reducing the calibration time t_j. If the desired number of supported instruments cannot be achieved by increasing calibration efficiency, we proceed to increment the number of ranks.

Once we have determined the required number of ranks in the scheme, we compute the actual necessary number of standards at each rank in the bottom–up manner, starting from N_{instr} and computing the number of the next rank up by a resolving (2.14) relative to N_j:

$$N_j = N_{j+1} \frac{t_{j+1}}{\eta_j T_{j+1}}. \tag{2.16}$$

Once we are done with the stem of the calibration scheme, we can add remaining branches one at a time as follows. Let j_{attach} be the rank of the lowest-accuracy secondary standards on the stem suitable to calibrate the instruments of the new branch, and $N_{j_{attach}+1}^{(\max)}$ be the maximum number of devices that could be serviced by standards at this rank according to (2.15). Then, $N^{(slack)} = N_{j_{attach}+1}^{(\max)} - N_{j_{attach}+1}$ gives the number of devices that could be added.

If the number of instruments at the new branch according to (2.16) does not exceed $N^{(slack)}$, we attach the new branch at rank j_{attach}, add the necessary number of standards at rank j_{attach}, and, moving from this rank up one step at a time, add the necessary number of standards at each rank (we are guaranteed that there will be enough capacity at each higher rank because the total number of devices at rank $j_{attach+1}$ does not exceed $N^{(\max)}_{j_{attach}+1}$).

Otherwise, that is, if the existing slack is insufficient, we must increase the capacity of the stem by adding an extra rank to add capacity. Accordingly, we recompute the number of devices at each rank of the stem in the bottom–up manner using (2.16), for the new number of ranks. After that, we repeat an attempt to attach the new branch from scratch.

If at some point we are unable to increment the number of ranks of the stem because the standard at the newly added rank would have insufficient accuracy, we would have to conclude that the given set of instruments is impossible to calibrate with the required accuracy using the available types of standards and the limitations on the use of the primary standard. However, given that the capacity of calibration schemes grows exponentially with the number of ranks, this outcome is practically impossible.

As the number of ranks increases, the capacity of the calibration network, represented by the checking scheme, increases rapidly. The calibration schemes in practice have at most five of ranks of standards, even for fields of measurement with large numbers of measuring instruments.

The relations presented above pertained to the simplest case, when at each step of transfer of the size of the unit, the period of time between calibrations and the calibration time were the same for all devices. In reality, these time intervals can be different for different types of devices. Taking this into account makes the calculations more complicated, but it does not change their essence. We consider these calculations next.

First, it is necessary to move from different time intervals between calibrations of different types of devices to one *virtual constant* time interval T_{vc} and to find the number of measuring instruments of each type N_k^{vc} that must be checked within this period. This is done using the obvious formula:

$$N_k^{vc} = N_k \frac{T_{vc}}{T_k}.$$

Next, it is necessary to find the average time t_j^{av} required to check one device for each step of the checking scheme:

$$t_j^{av} = \frac{\sum\limits_{k=1}^{n} t_k N_k^{vc}}{\sum\limits_{k=1}^{n} N_k^{vc}}. \tag{2.17}$$

Here n is the number of different types of devices at the jth step of the checking scheme.

We shall give a numerical example. Suppose it is required to organize a calibration of instruments of types A and B and the following data are given:

1. *Instruments of type* A: $N_{\mathrm{A}} = 3 \times 10^4$; the time interval between calibrations $T_{\mathrm{A1}} = 1$ year for $N_{\mathrm{A1}} = 2.5 \times 10^4$ and $T_{\mathrm{A2}} = 0.5$ year for $N_{\mathrm{A2}} = 5 \times 10^3$; the calibration time $t_{\mathrm{A}} = 5$ h.
2. *Instruments of type* B: $N_{\mathrm{B}} = 10^5$; $T_{\mathrm{B}} = 1$ year; the calibration time $t_{\mathrm{B}} = 2$ h.
3. *Primary measurement standard*: Four comparisons per year are permitted, and the utilization factor of the primary standard is $\eta_0 = 0.20$.
4. *Secondary standards:* the frequency of the calibration of secondary standards of rank 1 is 2 years; i.e., $T_1 = 2$ years; the time to perform one calibration is 60 h, and utilization factor $\eta_1 = 0.25$. For the devices of rank 2, $T_2 = 2$ years, $t_2 = 40$ h, and $\eta_2 = 0.25$. The calibration parameters of higher-rank standards are the same as those of the rank-2 standards.

The possible number of first-rank standards in this case is limited by the restrictions on the primary standards use and can be found as

$$N_1^{(\max)} = N_0 f T_1 = 8$$

because $N_0 = 1$; $f = 4$ is the maximum number of comparisons with a reference standard per year, and $T_1 = 2$. Obviously, eight standards are not enough to check 130,000 measuring instruments. We shall now see how many ranks of standards will be sufficient.

As the time between calibrations is different for different instruments, we pick the illusory constant time interval $T_{\mathrm{vc}} = 1$ year and find the number of instruments that must be checked within this time period. Conversion is necessary only for instruments of type A with $T_{\mathrm{A2}} = 0.5$ years, since the calibration interval of the rest of the instruments matches T_{vc}:

$$N_{A2}^{vc} = N_{A2} \frac{T_{vc}}{T_{A2}} = 5 \times 10^3 \times \frac{1}{0.5} = 10 \times 10^3$$

Therefore,

$$\sum_{k=A,B} N_k^{vc} = N_{AB} = N_{A1} + N_{A2}^{vc} + N_B = 135 \times 10^3$$

instruments must be calibrated within the time T_{ic}.

Different amounts of time are required to calibrate instruments of types A and B. The average calibration time t_{instr}^{av} of these working instruments, in accordance with (2.17), is

$$t_{instr}^{av} = \frac{(N_{A1} + N_{A2}^{vc})t_A + N_B t_B}{N_{AB}} = \frac{35 \times 10^3 \times 5 + 100 \times 10^3 \times 2}{135 \times 10^3} = 2.78\,\mathrm{h}.$$

Now, using (2.14), we shall find the maximum number of second-rank standards:

$$N_2^{(\max)} = N_1 \frac{\eta_1 T_2}{t_2} = 8 \times \frac{0.25 \times 2 \times 6 \times 10^3}{40} = 600.$$

The maximum number of instruments that can be calibrated with the above number of rank-2 secondary standards is

$$N_{instr}^{(\max)} = N_2^{(\max)} \frac{\eta_2 T_{vc}}{t_{instr}^{av}} = 600 \times \frac{0.25 \times 365 \times 24}{2.78} = 472661.$$

Here, $T_{\mathrm{vc}} = 365 \times 24 = 8.76 \times 10^3$ because 1 year = 365 days and η_2 was calculated for 24 h. The above number exceeds the total number of instruments N_{AB} to be calibrated; we thus conclude that two ranks are sufficient.

Next, we perform bottom–up calculations to find the necessary number of standards at each rank. The number of rank-2 standards is

$$N_2 = N_{AB} \frac{t_{instr}^{av}}{\eta_2 T_{vc}} = 135 \times 10^3 \times \frac{2.78}{0.25 \times 365 \times 24} = 171.$$

Similarly, one can check that all eight rank-1 secondary standards are needed, thus concluding the design of this calibration scheme.

Calculations similar to those in the above example allow one to choose in a well-grounded manner the structure of a calibration scheme and to estimate the required number of secondary standards of each rank. Calibration schemes in practice usually have extra capacity, which makes it possible to distribute secondary and working standards to limit their transport, to maximize the efficiency of calibration.

2.7 Statistical Analysis of Measuring Instrument Errors

A general characteristic of the errors of the entire population of measuring instruments of a specific type could be their distribution function. An important question then is if it is possible to find this function from experimental data. The studies in [47,54] have addressed this question using the data provided by calibration laboratories on instrument errors they observed during calibration. These data thus reflected the sample of instruments that were calibrated; because it is impossible to obtain the errors of all instruments of a given type that are in use, the use of a sampling method is unavoidable.

To establish a property of an entire group (general population) based on a sample, the sample must be representative. Sample homogeneity is a necessary indicator of representativeness. In the case of two samples, to be sure that the samples are homogeneous, it is necessary to check the hypothesis H_0: $F_1 = F_2$, where F_1 and F_2 are distribution functions corresponding, respectively, to the first and second sample.

The results of a calibration, as is well known, depend not only on the error of the measuring instrument being calibrated but also on the error of the standard. For this reason, measuring instruments calibrated with not less than a fivefold margin of accuracy (i.e., using a standard at least five times more accurate than the instrument) were selected for analysis.

In addition, to ensure that the samples are independent, they were formed either based on data provided by calibration laboratories in different regions of the former USSR or, in the case of a single laboratory, on the data separated by a significant time interval. The sample sizes were maintained approximately constant. Errors exceeding twice the limit of permissible error were deemed outliers and eliminated from the analysis.

The test of hypothesis H_0 was performed using the Wilcoxon and Siegel–Tukey criteria with a significance level $q = 0.05$. The technique of applying these criteria is described in Chap. 3. Table 2.2 shows the result of these tests obtained in the study of [47]. The table includes two samples, obtained at different times, for each instrument type. Rejection of the hypothesis is indicated by a minus sign, and acceptance is indicated by a plus sign. The symbol 0 means that a test based on the given criterion was not performed.

The Wilcoxon and Siegel–Tukey criteria are substantially different: The former is based on comparing averages, and the latter is based on comparing variances. For this reason, it is not surprising that there are cases when the hypothesis H_0 is rejected according to one criterion but accepted according to the other. The hypothesis of sample homogeneity must be rejected if even one of the criteria rejects it. Both samples of instruments of a given type were found to be homogeneous only for the Д566 wattmeters and standard manometers. For other measuring instruments, the compared samples were often found to be nonhomogeneous. It is

Table 2.2 The homogeneity hypothesis testing for samples of six types of measuring instruments

	Samples		Calibrated indication (each sample)	Result of hypothesis testing	
Instrument type	Year collected	Size		Wilcoxon	Siegel–Tukey
Э59 Ammeter	1974	160	30 divisions	+	−
			60	0	−
	1976	160	80	0	−
			100	+	+
Э59 Voltmeter	1974	120	70 divisions	−	0
	1976	108	150	+	+
Д566 Wattmeter	1974	86	70 divisions	+	+
	1976	83	150	+	+
TH-7 Thermometer	1975		100°C	0	−
			150°C	−	+
	1976		200°C	+	+
Standard spring manometer	1973	250			
			9.81 kPa	+	+
	1976	250			
P331 resistance measure	1970	400	10 kΩ	0	−
			100 Ω	0	−
	1975	400	10 Ω	0	−

interesting that the samples can be homogeneous on one scale marker, and inhomogeneous on another (see Э59 voltmeters and ammeters). TH-7 thermometers had homogeneous samples in one range of measurement and inhomogeneous samples in a different range. The calculations were repeated for significance levels of 0.01 and 0.1, but the results were generally the same in both cases.

The above experiment was formulated to check the stability of the distribution functions of the errors, but because the instruments in the compared samples were not always the same, the result obtained has a different but no less important meaning: It indicates that the samples are inhomogeneous. It means that the parameters of one sample are statistically not the same as these parameters of another sample of the same type of measuring instruments. Thus, the results obtained show that samples of measuring instruments are frequently nonhomogeneous with respect to errors. For this reason, they cannot be used to determine the distribution function of the errors of the corresponding instruments.

This result is also confirmed by the study of [54], which compared samples obtained from the data provided for Э59 ammeters by four calibration laboratories in different regions of the former USSR. The number of all samples was equal to 150–160 instruments. The errors were recorded at the markers 30, 60, 80, and 100 of the scale. The samples were assigned the numbers 1, 2, 3, and 4, and the hypotheses H_0: $F_1 = F_2$, $F_2 = F_3$, $F_3 = F_4$, and $F_4 = F_2$ were checked (the pairs of samples to compare were selected arbitrarily). The hypothesis testing was based on the Wilcoxon criterion with $q = 0.05$. The analysis showed that we can accept the hypothesis H_0: $F_1 = F_2$ only, and only at the marker 100. In all other cases, the hypothesis had to be rejected.

Thus, sampling does not permit finding the distribution function of the errors of measuring instruments. Moreover, the fact that the sampling data are unstable could mean that the distribution functions of the errors of the instruments change in time. There are definite reasons for this supposition.

Suppose that the errors of a set of measuring instruments of some type, at the moment they are manufactured, have a truncated normal distribution with zero mean. For measures (measuring resistors, shunts, weights, etc.), a measure with a too large positive error makes this measure impossible to repair (one could fix a weight whose mass exceeds the target by removing some material but one cannot repair a weight whose mass is too low). Furthermore, as measures age, their errors trend toward positive errors (e.g., weights lose some material due to polishing off with use). This is taken into account when manufacturing measures. For example, if in the process of manufacturing of a weight its mass is found to be even slightly less than the nominal mass then the weight is discarded. As a result, the distribution of the intrinsic errors of measures as they leave the factory is usually asymmetric.

Instrument errors change in the course of use. Usually the errors only increase. In those cases in which, as in the case of weights, the direction of the change of the errors is known beforehand and is taken into account during manufacturing, the errors can at first decrease, but then they will still increase. Correspondingly, changes in the instrument errors deform the distribution functions of the errors. This process, however, does not occur only spontaneously. At the time of routine checks, measuring

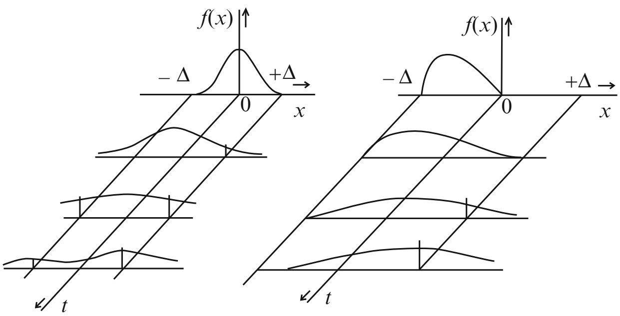


Fig. 2.6 Examples of possible changes in the probability densities of the errors of measuring devices in time. The figure on the left shows an example of changes in error distribution of a batch of measurement instruments; the figure on the right shows a possible change in error distribution of a batch of weights

instruments whose errors exceed the established limits are discarded, which again affects the distribution function of the errors of the remaining instruments.

The right-hand side of Fig. 2.6 shows the approximate qualitative picture of the changes occurring in the probability distribution of errors of a batch of weights in time. It shows the initial distribution of errors with all the errors being negative. With time, as the measures wear off, their errors decrease, with some positive errors starting to appear. As this trend continues, at some point some instruments start being discarded (which is shown in the figure by a vertical cut-off line at $+\Delta$ error limit). The process ultimately terminates when the measuring instruments under study no longer exist: either their errors exceed the established limits or they are no longer serviceable for other reasons.

The left-hand side of this figure shows an example of changes in error distribution in a batch of measuring instruments. In this example, the errors generally increase in time but the change is biased toward positive errors. Again, at some point instruments start to be discarded, but most of the discarded instruments are those with positive errors.

There are other evident reasons for this result. One reason is that the stock of instruments of each type is not constant. On the one hand, new instruments that have just been manufactured are added to the stock. On the other hand, in the verification, some instruments are rejected, and some instruments are replaced. The ratio of the numbers of old and new instruments is constantly changing. Another reason is that groups of instruments are often used under different conditions, and the conditions of use affect differently the rate at which the instrumental errors change.

The temporal instability of measuring instruments raises the question of whether the errors of measuring instruments are in general sufficiently stable so that a collection of measuring instruments can be described by some distribution function. At a fixed moment in time, each type of instruments without doubt can be described by distribution function of errors. But the problem is how to find this distribution function. The simple sampling method, as we saw above, is not suitable. Moreover, even if the distribution function could be found by some complicated method, after some time, it would have to be redetermined, because the errors, and the composition of the stock of measuring instruments, change. Therefore, we have to conclude that the distribution of errors of measuring instruments cannot be found based on the experimental data.

The results presented above were obtained in the former USSR, and instruments manufactured in the former USSR were studied. However, there is no reason to expect that instruments manufactured in other countries will have different statistical properties.

Chapter 3
Statistical Methods for Experimental Data Processing

3.1 Methods for Describing Random Quantities

The presence of random errors in measurements leads to the wide usage of the concept of *random quantity* as a mathematical model for random errors and, equivalently, for measurement results. The realization of the random error in a given act of measurement is called the random error of a separate measurement, and the word "separate" is often omitted for brevity. Where it can cause confusion between a separate measurement and a complete measurement (which may comprise multiple separate measurements), we will refer to the results of separate measurements as *observations*.

Random quantities are studied in the theory of probability, a well-developed field of mathematics. The properties of a random quantity are completely described by the distribution function $F(x)$, which determines the probability that a random quantity X will assume a value less than x:

$$F(x) = P\{X < x\}.$$

The distribution function is a nondecreasing function, defined so that $F(-\infty) = 0$ and $F(+\infty) = 1$. It is said to be cumulative or integral.

Continuous and discrete random variables are distinguished. For continuous random variables, together with the cumulative distribution function $F(x)$, the differential function, usually called the probability density $f(x)$, is also widely employed:

$$f(x) = \frac{dF(x)}{dx}.$$

We call attention to the fact that the probability density is a dimensional function:

$$\dim f(x) = \dim \frac{1}{X}.$$

In the practice of precise measurements one most often deals with normal and uniform distributions. Figure 3.1a shows integral functions of these distributions, and Fig. 3.1b shows the probability densities of the same distributions.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_3,

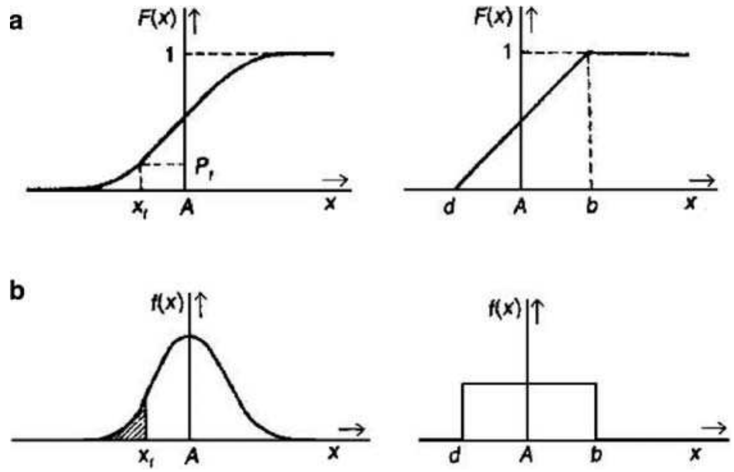


Fig. 3.1 (**a**) The probability distribution and (**b**) the probability density for a normal distribution (on the *left*) and uniform distribution (on the *right*) of continuous random quantities

For the normal distribution, we have

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-A)^2/2\sigma^2},$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-A)^2}/2\sigma^2\, dx, \tag{3.1}$$

The parameter σ^2 is the variance, and A is the mathematical expectation of the random quantity. A normal distribution is fully determined by its mathematical expectation and variance, and is often denoted as $N(A, \sigma^2)$.

The value of $F(x)$ for some fixed x_f gives the probability $P\{X < x_f\} = P_f$.

When the graph of $f(x)$ is used to calculate this probability, it is necessary to find the area under the curve to the left of the point x_f. The left side of Fig. 3.1 illustrates finding P_f from cumulative distribution and density functions.

To avoid tabulating functions (3.1) for every specific values of σ and A, calculations widely rely on the *standard* normal distribution, which is obtained by transforming the random quantity X to $Z = (X - A)/\sigma$. Random variable Z is normally distributed with mathematical expectation 0 and variance 1. Its probability distribution and density functions are:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{2} e^{-y^2/2} dy. \tag{3.2}$$

Customarily, calculations related to normal distribution are based on the function $\Phi(z)$ below, instead of (3.2):

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-y^2/2} dy \tag{3.3}$$

Function $\Phi(z)$ is called the standard Gaussian function, and its values are given in Table A.1 in the Appendix.

It is obvious that for $z \geq 0$

$$F(z) = 0.5 + \Phi(z).$$

The branch for $z < 0$ is found based on symmetry considerations:

$$F(z) = 0.5 - \Phi(z).$$

The normal distribution is remarkable in that according to the central limit theorem, the sum of a number of random quantities with arbitrary distributions tends to a normal distribution as the number of random quantities grows to infinity. In practice, the distribution of the sum of a comparatively small number of random quantities already is found to be close to a normal distribution.

The uniform distribution is defined as

$$f(x) = \begin{cases} 0, & x < d, \\ \frac{1}{b-d}, & d \geq x \leq b, \\ 0, & x > b, \end{cases}$$

$$F(x) = \begin{cases} 0, & x < d, \\ \frac{x-d}{b-d}, & d \geq x \leq b, \\ 1, & x > b. \end{cases} \tag{3.4}$$

We shall also use the uniform distribution often.

In addition to continuous random variables, discrete random variables are also encountered in metrology. An example of an integral distribution function and the probability density of a discrete random variable are given in Fig. 3.2.

Distribution functions are complete characteristics of random quantities, but they are not always convenient to use in practice. For this reason, random quantities are also characterized by their numerical parameters called moments. The initial moments m_k (moments about zero) and central moments μ_k (moments about the mean value) of order k are defined by the formulas

$$m_k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx,$$

$$m_k = E[X^k] = \sum_{i=1}^{n} x_i^k P_i. \tag{3.5}$$

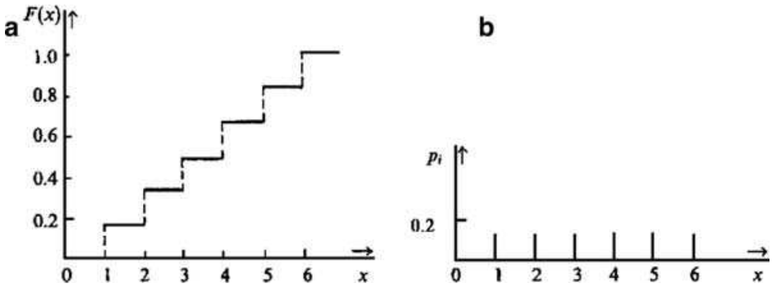


Fig. 3.2 (**a**) The probability distribution and (**b**) the probability density of a discrete random quantity

$$\mu_k = E[X - E[x]]^k = \int_{-\infty}^{\infty} (x - E[X])^k f(x)dx,$$
$$\mu_k = E[X - E[x]]^k = \sum_{i=1}^{n} (x_i - E[X])^k p_i. \tag{3.6}$$

In the relations (3.5)–(3.8), the first formulas refer to continuous and the second to discrete random quantities.

Of the initial moments, the first moment ($k = 1$) is most often employed. It gives the mathematical expectation of the random quantity

$$m_1 = E[X] = \int_{-\infty}^{\infty} x f(x)dx,$$
$$m_1 = E[X] = \sum_{i=1}^{n} x_i p_i. \tag{3.7}$$

It is assumed that $\sum_{i=1}^{n} p_i = 1$; i.e., the complete group of events is considered.

Of the central moments, the second moment ($k = 2$) plays an especially important role. It is the variance of the random quantity

$$\mu_2 = \mathrm{V}[X] = E[(X - m_1)^2] = \int_{-\infty}^{\infty} (x - m_1)^2 f(x)dx,$$
$$\mu_2 = \mathrm{V}[X] = E[(X - m_1)^2] = \sum_{i=1}^{n} (x_i - m_1)^2 p_i. \tag{3.8}$$

The square root of the variance is called the standard deviation of the random quantity

$$\sigma = +\sqrt{V[X]} \tag{3.9}$$

Correspondingly, $V[X] = \sigma^2$.

The third and fourth central moments are also used in applications. They are used to characterize the symmetry and sharpness of distributions. The symmetry is characterized by the skewness $a = \mu_3/\sigma^3$, and the sharpness is characterized by the excess $e = \mu_4/\sigma^4$. The latter is sometimes defined as $e' = \mu_4/\sigma^4 - 3$ because normal distribution has $e = 3$.

The normal distribution is completely characterized by two parameters: $m_1 = A$ and σ. For it, characteristically, $a = 0$ and $e' = 0$. The uniform distribution is also determined by two parameters: $m_1 = A$ and $l = d - b$. It is well known that

$$m_1 = \frac{d+b}{2}, \qquad V[X] = \frac{(d-b)^2}{12} = \frac{l^2}{12}. \tag{3.10}$$

Instead of l, the quantity $h = l/2$ is often used. Then $V[X] = h^2/3$ and $\sigma(X) = h/\sqrt{3}$.

3.2 Requirements for Statistical Estimates

As mentioned in the previous section, the probability distribution function and the probability density fully describe the properties of a random quantity. Unfortunately, they are rarely available. Consequently, one has to estimate parameters of a random quantity from statistical data, that is, from the observations of the random quantity.

Given a specific sample of observations, any estimate derived from this sample is a specific number. However, across different samples, the estimate will be different, and it is a random variable for a random sample. Thus, one can talk about statistical properties of the estimates.

The estimates obtained from statistical data must be consistent, unbiased, and efficient.

An estimate $\tilde{A}$ is said to be consistent if, as the number of observations increases, it approaches the true value of the estimated quantity A (it converges probabilistically to A):

$$\tilde{A}(x_1, \ldots, x_n) \underset{n \to \infty}{\to} A.$$

The estimate of A is said to be unbiased if its mathematical expectation is equal to the true value of the estimated quantity:

$$E[\tilde{A}] = A.$$

In the case when several unbiased estimates can be found, the estimate that has the smallest variance is, naturally, regarded as the best estimate. The smaller the variance of an estimate the more efficient it is.

Methods for finding estimates of a measured quantity and indicators of the quality of the estimates depend on the form of the distribution function of the observations. For a normal distribution of the observations, the arithmetic mean of the observations, as well as their median (which is the point x_m such that

$P\{X < x_m\} = P\{X > x_m\}$) can be taken as an estimate of the true value of the measured quantity. The ratio of the variances of these estimates is well known [19]:

$$\sigma_{\bar{x}}^2/\sigma_m^2 = 0.64,$$

where $\sigma_{\bar{x}}^2$ is the variance of the arithmetic mean and σ_m^2 is the variance of the median. Therefore, the arithmetic mean is a more efficient estimate of A than the median.

In the case of a uniform distribution, the arithmetic mean of the observations or the half-sum of the minimum and maximum values can be taken as an estimate of A:

$$\tilde{A}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \tilde{A}_2 = \frac{x_{\min} + x_{\max}}{2}.$$

The ratio of the variances of these estimates is also well known [19]:

$$\frac{V[\tilde{A}_1]}{V[\tilde{A}_2]} = \frac{(n+1)(n+2)}{6n}.$$

For $n = 2$, this ratio is equal to unity, and it increases for $n > 2$. For example, for $n = 10$, it is already equal to (2.2), making the half-sum of the minimum and maximum values in this case a more efficient estimate than the arithmetic mean.

3.3 Evaluation of the Parameters of the Normal Distribution

If the available data are consistent with the hypothesis that the observations belong to a normal distribution, then it is sufficient to estimate the expectation $E[X] = A$ and the variance σ^2 to describe fully the distribution. We will discuss methods of obtaining these estimates in this section.

When the probability density of a random quantity is known, its parameters can be estimated by the method of maximum likelihood. We shall use this method to find the estimates above.

The elementary probability of obtaining some specific observation x_i in the interval $x_i \pm \Delta x_i/2$ is equal to $f_i(x_i, A, \sigma)\Delta x_i$, where f_i is the value of the probability density function with parameters A and the σ for point x_i. Assume that all observations are independent. Then, the probability of encountering all experimentally obtained observations with $\Delta x_1, \cdots, \Delta x_n$ is equal to

$$P_l = \prod_{i=1}^{n} f_i(x_i, A, \sigma)\Delta x_1 \cdots \Delta x_n.$$

The idea of the method is to take for the estimate of the parameters of the distribution (in our case, A and σ), the values that maximize the probability P_l. These values are found, as usual, by equating to zero the partial derivatives of P_l with respect to the parameters being estimated. The constant cofactors do not affect the solution, and for this reason, only the product of the functions f_i is considered; this product is called the likelihood function:

$$L(x_1, \ldots, x_n; A, \sigma) = \prod_{i=1}^{n} f_i(x_1, \ldots, x_n; A, \sigma).$$

We now return to our problem. For the available group of observations $x_1, \ldots, x_n$, the values of the probability density will be

$$f_i(x_i, A, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i - A)^2/2\sigma^2}.$$

Therefore,

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(xi - A)^2\right).$$

To find the maximum of L, it is convenient to investigate $\ln L$:

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - A)^2.$$

The maximum of L will occur when $\partial L/\partial A = 0$ and $\partial L/\partial\sigma^2 = 0$:

$$\frac{\partial L}{L\partial A} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - A) = 0,$$

$$\frac{\partial L}{L\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - A)^2 = 0.$$

From the first equation, we find an estimate for A:

$$\tilde{A} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.11}$$

The second equation gives the estimate $\tilde{\sigma}^2 = (1/n)\sum_{i=1}^{n}(x_i - A)^2$. But A is unknown; taking instead of A its estimate $\bar{x}$, we obtain

$$\tilde{\sigma}_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Let us now check to see whether the obtained estimates are consistent and unbiased. Because all x_i are drawn from the same distribution, the mathematical expectation of the ith observation in a random sample is equal to A for every i: $E(x_i) = A$.[1] For this reason,

$$E[\tilde{A}] = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = A.$$

Therefore, $\tilde{A}$ is an unbiased estimate of A. It is also a consistent estimate, because as $n \to \infty$, $\tilde{A} \to A$, according to the law of large numbers.

We shall now investigate $\tilde{\sigma}_*^2$. In the formula derived above, the random quantities are x_i and $\bar{x}$. For this reason, we shall rewrite it as follows:

$$\begin{aligned}
\tilde{\sigma}_*^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - A + A - \bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}[(x_i - A)^2 - 2(x_i - A)(\bar{x} - A) + (\bar{x} - A)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 - \frac{2}{n}\sum_{i=1}^{n}(x_i - A)(\bar{x} - A) + \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - A)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 - (\bar{x} - A)^2,
\end{aligned}$$

because

$$\frac{1}{n}\sum_{i=1}^{n}(\bar{x} - A)^2 = (\bar{x} - A)^2$$

and

$$\frac{2}{n}\sum_{i=1}^{n}(x_i - A)(\bar{x} - A) = \frac{2}{n}(\bar{x} - A)\sum_{i=1}^{n}(x_i - A) = 2(\bar{x} - A)^2.$$

We shall find $E[\tilde{\sigma}_*^2]$. To this goal, the following relations must be used. By definition, according to (3.8), we have $E(x_i - A)^2 = \sigma^2$. Therefore,

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(x_i - A)^2\right] = \sigma^2.$$

For the random quantity $\bar{x}$, we can write analogously $E(\bar{x} - A)^2 = V[\bar{x}]$. We can express $V[\bar{x}]$ in terms of σ^2 as follows

$$V[\bar{x}] = V\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n}V[X] = \frac{\sigma^2}{n}.$$

[1] With a slight abuse of notation, we use x_i to denote the ith observation in both a specific sample (where it is just a number) and in a random sample (where it is a random variable).

Thus,

$$E[\tilde{\sigma}_*^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Therefore, the obtained estimate $\tilde{\sigma}_*^2$ is biased. But as $n \to \infty$, $E[\tilde{\sigma}_*^2] \to \sigma^2$, and therefore, this estimate is consistent.

To correct the estimate, i.e., to make it unbiased, $\tilde{\sigma}_*^2$ must be multiplied by the correction factor $n/(n-1)$. Then we obtain

$$\tilde{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3.12}$$

This estimate is also consistent, but, as one can easily check, it is now unbiased. Some deviation from the maximum of the likelihood function is less important for us than the biasness of the estimate.

The standard deviation of the random quantity X is $\sigma = \sqrt{V[X]}$, and it is not a random quantity. Instead of σ^2 we must use the estimate of the variance from (3.12) – a random quantity. Extracting the square root is a nonlinear procedure; it introduces bias into the estimate $\tilde{\sigma}$. To correct this estimate, a factor k_n, depending on n as follows, is introduced:

N	3	4	5	6	7	10
k_n	1.13	1.08	1.06	1.05	1.04	1.03

So,

$$\tilde{\sigma} = k_n\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{3.13}$$

The following formula gives approximately the same result [28]:

$$\tilde{\sigma} = \sqrt{\frac{1}{n-1.5}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

As the number of observations is rarely large, the error in the determination of the standard deviation can be significant. In particular, this error is typically significantly larger than the biasness introduced into the estimate by the square root extraction. For this reason, in practice, the latter can usually be neglected and the correction factor k_n not employed. Thus, instead of (3.13), the estimate of the standard deviation is commonly found as the square root of the variance given by (3.12). Therefore, the estimate of the standard deviation is calculated as follows:

$$\tilde{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}. \tag{3.14}$$

We have obtained estimates of the parameters of the normal distribution, but they are also random quantities: When the measurement is repeated, we obtain a different group of observations with different values of $\bar{x}$ and $\tilde{\sigma}$. The spread in these estimates can be characterized by their standard deviations $\sigma(\bar{x})$ and $\sigma(\tilde{\sigma})$. We already obtained above that $V[\bar{x}] = \sigma^2/n$. Therefore,

$$\sigma(\bar{x}) = \sqrt{V[\bar{x}]} = \frac{\sigma}{\sqrt{n}}. \tag{3.15}$$

By replacing σ in (3.15) with its estimate from (3.14), we can obtain an estimate of $\sigma(\bar{x})$, denoted as $\tilde{\sigma}(\bar{x})$ or, more commonly, $S_{\bar{x}}$ or $S(\bar{x})$:

$$S(\bar{x}) = \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}. \tag{3.16}$$

Uncertainty of the estimate given in (3.16) depends on the number of measurements n and on the confidence probability α. The method of computing this uncertainty is described in Sect. 3.5.

This uncertainty can be sizable; for example, for $n = 25$ and $\alpha = 0.80$, the uncertainty of this estimate is about 20%; for $n = 15$ and $\alpha = 0.80$, it is about 30%. However, this uncertainty is not taken into account when estimating the uncertainty of the measurement result. It would be interesting to understand why this does not cause problems in practice.

3.4 Elimination of Outlying Data

If in the group of observations, one or two differ sharply from the rest, and no slips of the pen, reading errors, and similar blunders have been found, then it is necessary to decide whether they are extreme events that should be excluded. This problem is solved by statistical methods based on the assumption that the distribution from which the observations are drawn is normal. The methodology for solving the problem is presented in the standard [4].

The solution scheme is as follows. An ordered series $x_1 < x_2 < \cdots < x_n$ is constructed from the obtained observations. The candidate to be tested for outlier is obviously x_1 or x_n. From all x_i, we calculate $\bar{x}$ and, using (3.14), the estimate of the standard deviation of this group of observations, S. We next compute how much the potential outlier candidate deviates from the mean value:

$$t_1 = \frac{\bar{x} - x_1}{S} \tag{3.17}$$

and

$$t_n = \frac{x_n - \bar{x}}{S}. \tag{3.18}$$

Now we select the candidate to be tested that has the bigger deviation among the two above. Let us assume that it is x_1. We resort to the Table A.3 reproduced in Appendix, which is read as follows. For a given number of observations n and chosen percentage q (referred to as significance level) and corresponding value T, q is the probability that t_1 exceeds T.

In other words, if the value of t_1 is greater than T for a selected significance level, then the corresponding value of x_1 can be discarded: The probability that a "legitimate" observation (i.e., an observation belonging to the distribution) would produce $t > T$ is less than or equal to the adopted significance level. Thus, the significance level gives the probability that we erroneously discard an observation that in fact belongs to the distribution.

If we want to estimate probability of encountering an outlier in a future similar measurement, we must take into account that the outlier can be either too big or too small. Either observation can occur with an equal probability, due to the symmetry of the normal distribution. Thus, the probability of encountering either of them is equal to $2q$.

The described procedure is quite useful and is widely employed in statistical data processing. But one could say that an "abnormal" observation may actually reflect some unknown feature of the subject under study and thus should not be discarded lightly. Let us consider this issue in more detail.

Imagine a measurement in which an observation occurred that seems atypically different from others. What will an expert performing this measurement do? First, he or she will check if any physical properties of the object under study, or any other physical reasons, might have caused the unusual observation. If this check does not lead to an explanation for this observation, the expert will analyze all the aspects of the measurement procedure, measurement conditions, and records documenting the measurement execution. If there is still no rational explanation for the unusual observation, the expert will conduct a statistical analysis using methods described earlier in this section, to check if this observation could be an outlier. If this analysis confirms that the observation is an outlier, it can be discarded. However, in especially important cases, such as when the decision can affect public safety, the expert may chose to continue the experiment collecting more observations. More observations may reveal physical or other reasons behind the abnormality. If not, the expert will repeat the statistical analysis, this time using all the accumulated data, and based on its result, will make the final decision on accepting or discarding the observation. When will the expert stop collecting more data? Only his or her experience and intuition will tell. Unfortunately, there is no prescribed procedure here to follow. However, there are the following two general reasons to discard the observation detected as an outlier by statistical analysis:

1. A real measurement as a rule consists of a small number of observations, and the probability of them including more than one outlier is extremely small. Therefore, this outlier cannot be compensated with another one having the opposite sign.

2. Because the outlier deviates significantly from the rest of the results, it skews the average value of the set of data. In other words, not only does it increase the inaccuracy of a measurement, but also affects the measurement result.

Thus, if there are no physical reasons for the outlying result, it must be discarded.

Example 3.1. Assume ten repeated measurements of the current strength in mA gave the following results: 10.07, 10.08, 10.10, 10.12, 10.13, 10.15, 10.16, 10.17, 10.20, and 10.40. The value 10.40 differs sharply from the other values. We shall check to see whether or not it can be discarded. We shall use the criterion presented, though we do not have the data that would allow us to assume that these observations satisfy the normal distribution.

The mean and standard deviation of this group of observations are $\bar{x} = 10.16$ mA and $S = 0.094$ mA. According to the procedure, we compute

$$t_{10} = \frac{(10.40 - 10.16)}{0.094} = 2.55.$$

Let us select significance level of 0.5%. Turning to Table A.3, we find critical value T for $n = 10$ and $q = 0.5\%$. This value is $T = 2.48$. Since $t_{10} > T$, we conclude that assuming this observation to be an outlier would be incorrect only with probability at most 0.5%.

3.5 Construction of Confidence Intervals

Having obtained the estimate $\tilde{A}$, it is of interest to determine by how much it can change in repeated measurements performed under the same conditions. This question is clarified by constructing the *confidence interval* for the true value of the measured quantity.

The *confidence interval* is the interval that includes, with a prescribed probability called the *confidence probability*, the true value of the measurand. The concepts of confidence interval and confidence probability can be interpreted as follows. Imagine a quantity that is measured multiple times under the same conditions, where each measurement can itself comprise multiple observations. Assume that we use the data obtained from each of these measurements to build the confidence interval corresponding to the same confidence probability 0.95. Then, 95% of the obtained confidence intervals will cover the true value of the measured quantity.

Confidence intervals are often expressed as $(x \pm \Delta x)$ or $(x \pm \delta\%)$, where x is the center of the interval and Δx and $\delta\%$ represent the half-length of the interval in the absolute or relative form. The latter values define the limits of the confidence interval. We will, therefore, refer to the half-length of the confidence interval as the *confidence limit*.

In principle, the confidence interval could be constructed based on the Chebyshev's inequality [19]:

$$P\{|X - A| \geq t\sigma\} \leq \frac{1}{t^2}$$

where t is a parameter dependent on the confidence probability, which will be explained shortly.

For the random quantity $\bar{x}$, we obtain, using (3.15):

$$P\left\{|\bar{x} - A| \geq \frac{t\sigma}{\sqrt{n}}\right\} \leq \frac{1}{t^2}. \tag{3.19}$$

Let us transform the inequality (3.19) so that it would determine the probability that a deviation of the random quantity from its true value is less than a certain value. After simple transformations, we obtain

$$P\left\{|\bar{x} - A| \leq t\frac{\sigma}{\sqrt{n}}\right\} \geq 1 - \frac{1}{t^2}.$$

Without knowing anything about the distribution of the random errors, the coefficient t can be calculated based on a prescribed confidence probability α from the right-hand side of the above inequality, which gives

$$t = \frac{1}{\sqrt{1-\alpha}}.$$

Then, the confidence interval for α follows from the above inequality and is:

$$\left[\bar{x} - t\frac{\sigma}{\sqrt{n}}, \bar{x} + t\frac{\sigma}{\sqrt{n}}\right].$$

If the distribution of the random errors can be assumed to be symmetric relative to A, then the confidence interval can be narrowed somewhat [19], using the inequality

$$P\left\{|\bar{x} - A| \leq t\frac{\sigma}{\sqrt{n}}\right\} \geq 1 - \frac{4}{9}\frac{1}{t^2}.$$

where

$$t = \frac{2}{3\sqrt{1-\alpha}}.$$

In either case, the standard deviation of the results of measurements σ can be estimated with (3.16) and then the confidence interval can be found.

Using Chebyshev's inequality is attractive because it does not require one to know the form of the distribution function of the observations. It uses the arithmetic mean as the estimation of the measured quantity, which can practically always be done (although in the case when the distribution differs from a normal distribution, the estimate will not be the most efficient estimate). However, the confidence intervals constructed in this manner are only approximate, because the effect of replacing the standard deviation by its estimate is not taken into account. More importantly, the intervals obtained with the help of the Chebyshev's inequality are too wide for practice, and so this method is rarely (if ever) used.

If the distribution of the observations can be regarded as normal with a known standard deviation, then the confidence interval is constructed based on the expression

$$P\left\{|\bar{x} - A| \le z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\} = \alpha.$$

where α is the selected confidence probability and $z_{\frac{1+\alpha}{2}}$ is the quantile of the standard normal distribution for probability $\frac{1+\alpha}{2}$. (By the quantile of a distribution with cumulative distribution function F for probability p we mean the value x such that $F(x) = p$).

For example, let $\alpha = 0.95$. With this probability, the interval

$$\left\lfloor \bar{x} - z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\rfloor$$

should include the true value of A. The probability that A falls outside this interval is equal to $1 - \alpha = 0.05$. As the normal distribution is symmetric, the probabilities that A falls beyond either limit of the interval are the same and equal to $(1-\alpha)/2 = 0.025$. It is obvious that the cumulative probability of the upper limit of this interval is $(1 - 0.025) = 0.975$. It can be calculated as

$$P = 1 - \frac{1-\alpha}{2} = \frac{1+\alpha}{2}.$$

We shall now show how to find the value of $z_{\frac{1+\alpha}{2}}$, using the standard Gaussian function, whose values are given in Table A.1 of the Appendix. The standard Gaussian function $\Phi(z)$ is related to the standard normal distribution function $F(z)$ by the relation $F(z) = 0.5 + \Phi(z)$, or $\Phi(z) = F(z) - 0.5$. Therefore, the quantile of $F(z)$ for probability $\frac{1+\alpha}{2} = 0.975$ is the same as the quantile of $\Phi(z)$ for probability $0.975 - 0.5 = 0.475$. Using Table A.1, we find the quantile $z_{0.975} = 1.96$ corresponding to the argument 0.475.

Often the value of the quantile $z_{\frac{1+\alpha}{2}}$ is given and the corresponding probability α needs to be found. For example, for $z_{\frac{1+\alpha}{2}} = 1$, $\Phi(z) = 0.3413$, and $F(z) = \Phi(z) + 0.5 = 0.841$. Then $(1 + \alpha)/2 = 1 - F^2(z) = 0.159$ and $\alpha = 0.682$. Analogously, for $z_{\frac{1+\alpha}{2}} = 3$, we find $\Phi(z) = 0.49865$, $F(z) = 0.99865$, $(1 + \alpha)/2 = 0.00135$, and $\alpha = 0.9973$.

So far we explained how we could build the confidence interval from the quantile $z_{\frac{1+\alpha}{2}}$ assuming we know the standard deviation σ. In practice, however, the standard deviation is rarely known. Usually we know only its estimate S and, correspondingly, $S_{\bar{x}} = S/\sqrt{n}$. Then, still assuming that the observations can be viewed as belonging to a normal distribution, the confidence intervals are constructed based on Student's t distribution. The applicability of Student's distribution is based on the property that if a random quantity x is normally distributed, then the random quantity

$$t = \frac{\bar{x} - A}{S_{\bar{x}}},$$

obtained from random samples of size n, belongs to Student's distribution with $(n-1)$ degrees of freedom. In the above formula, $S_{\bar{x}}$ is the estimate of the standard deviation of the arithmetic mean $\bar{x}$, calculated from (3.16). Then, the confidence interval $\left[\bar{x} - t_q S_{\bar{x}}, \bar{x} + t_q S_{\bar{x}}\right]$ corresponds to the probability

$$\boldsymbol{P}\{|\bar{x} - A| \le t_q S_{\bar{x}}\} = \alpha,$$

where t_q is the q th percentile of Student's distribution with the degrees of freedom $\nu = n - 1$. Traditionally, tables for Student's distribution list percentiles for probability function $P\{t > t_q\}$. We present such a table as Table A.2 in Appendix. Thus, given α, we obtain the *significance level* $q = 1 - \alpha$, then look up t_q in Table A.2 for this significance level and the degrees of freedom $\nu = n - 1$, and finally compute the confidence interval above that corresponds to α. The confidence interval is commonly represented by confidence limits:

$$u = t_q S_{\bar{x}}. \tag{3.20}$$

In measurement practice, the confidence probability is increasingly often set to 0.95. Further, confidence intervals are in practice constructed almost always based on Student's distribution as just described. This method is widely applicable because experimental data are typically symmetrical around the mean, and in this case, this method is used even when the distribution of the underlying random quantity x deviates from normal. Indeed, as seen from (3.20), Student's distribution is determined by $\bar{x}$ and $S_{\bar{x}}$, and is not directly dependent on x and therefore is robust.

Sometimes confidence intervals are constructed for the standard deviation. In these cases, the χ^2 distribution is employed. This method relies on the property that if a random quantity x is normally distributed, then the random quantity

$$\chi^2 = \frac{(n-1)\tilde{\sigma}^2}{\sigma^2},$$

obtained from random samples of size n, belong to χ^2 distribution with n degrees of freedom. Unlike Student's distribution, χ^2 distribution is asymmetrical, and we must use different quantiles to compute lower and upper limits of the confidence interval. Consequently, the confidence interval for the confidence probability is

$$P\left\{\left(\frac{\sqrt{n-1}}{\chi_L}\right)\tilde{\sigma} \le \sigma \le \left(\frac{\sqrt{n-1}}{\chi_U}\right)\tilde{\sigma}\right\} = \alpha \tag{3.21}$$

is found as follows. Table A.4 gives percentiles of the probability function $P\left\{\chi^2 > \chi_q^2\right\}$. Given confidence probability α, we find the probabilities corresponding to the lower and upper limits of the confidence interval: $p_L = (1-\alpha)/2$ and $p_U = (1+\alpha)/2$. We then, conceptually, obtain significance levels $q_L = 1 - P_L$ and $q_U = 1 - P_U$. Next, from Table A.4, we look up the p_Lth and p_Uth percentiles (denote them, respectively, as χ_L^2 and χ_U^2) for the probability function $P\left\{\chi^2 > \chi_q^2\right\}$. Again, we use the degree of freedom $\nu - n - 1$ because there is an unknown quantity σ^2 in the expression for χ^2. Finally, we use χ_L^2 and χ_U^2 to compute the confidence interval for σ. Because σ has inverse dependence on χ, χ_L^2 determines the upper limit of the confidence interval and χ_U^2 the lower limit.

For example, let $\tilde{\sigma} = 1.2 \times 10^{-5}$ and $n = 10$. Take $\alpha = 0.90$. Then $p_U = (1+0.9)/2 = 0.95$ and $p_L = (1-0.9)/2 = 0.05$. The degree of freedom $\nu = 10 - 1 = 9$. From Table A.4, we find $\chi_U^2 = 3.325$ and $\chi_L^2 = 16.92$. The confidence interval will then be

$$\left[\frac{\sqrt{10-1}}{\sqrt{16.92}} \times 1.2 \times 20^{-5}, \frac{\sqrt{10-1}}{\sqrt{3.325}} \times 1.2 \times 10^{-5}\right];$$

i.e.,

$$[0.88 \times 10^{-5} \le \sigma \le 2.0 \times 10^{-5}].$$

When constructing confidence intervals for standard deviation, the confidence probability can be taken to be less than the confidence probability in the case of constructing the confidence interval for the true value of the measured quantity. Often $\alpha = 0.80$ is sufficient.

Confidence intervals should not be confused with *statistical tolerance intervals* (first mentioned at the end of Sect. 2.3). The statistical tolerance interval is the interval that, with prescribed probability a, contains not less than a prescribed fraction p_0 of the entire collection of values of the random quantity (population). Thus, the statistical tolerance interval is the interval for a random quantity, and this distinguishes it principally from the confidence interval that is constructed to cover the value of a nonrandom quantity.

If, for example, the sensitivity of a group of strain gauges is measured, then the obtained data can be used to find the interval with limits l_1 and l_2 in which, with prescribed probability a, the sensitivity of not less than the fraction p_0 of the entire batch (or the entire collection) of strain gauges of the given type will fail. This is the statistical tolerance interval. Methods for constructing this tolerance interval can be found in books on the theory of probability and mathematical statistics.

One must also guard against confusing the limits of statistical tolerance and confidence intervals with the tolerance range for the size of some parameter. The tolerance or the limits of the tolerance range are, as a rule, determined before the fabrication of a manufactured object, so that the objects for which the value of the parameter of interest falls outside the tolerance range are unacceptable and are discarded. In other words, the limits of the tolerance range are strict limits that are not associated with any probabilistic relations.

The statistical tolerance interval, however, is determined by objects that have already been manufactured, and its limits are calculated so that with a prescribed probability, the parameters of a prescribed fraction of all possible manufactured objects fall within this interval. Thus, the limits of the statistical tolerance interval, as also the limits of the confidence interval, are random quantities, whereas the tolerance limits or tolerances are nonrandom quantities.

3.6 Testing Hypotheses about the Form of the Distribution Function

The problem is usually posed as follows: For a group of measurement results, it is hypothesized that these results can be regarded as realizations of a random quantity with a distribution function having a chosen form. Then this hypothesis is checked by the methods of mathematical statistics and is either accepted or rejected.

For a large number of observations ($n > 50$), Pearson's test (χ^2 test) for grouped observations and the Kolmogorov–Smirnov test for nongrouped observations are regarded as the best tests. These methods are described in many books devoted to the theory of probabilities and statistics. For example, see [19, 49, 53]. We shall discuss the χ^2 test, and for definiteness, we shall check the data on belonging to a normal distribution.

The idea of this method is to monitor the deviations of the histogram of the experimental data from the histogram with the same number of intervals that is constructed based on the normal distribution. The sum of the squares of the differences of the frequencies over the intervals must not exceed the values of χ^2 for which tables were constructed as a function of the significance level of the test q and the degree of freedom $\nu = L - 3$, where L is the number of intervals and minus 3 is because the measurement data have two unknown parameters (the mathematical expectation and variance) and χ^2 distribution has one more unknown parameter (its degree of freedom).

The calculations are performed as follows:

1. The arithmetic mean of the observations and an estimate of the standard deviations are calculated.
2. Measurements are grouped according to intervals. For about 100 measurements, five to nine intervals are normally taken. For each interval, the number of measurements $\tilde{\varphi}_i$ falling within the interval is calculated.
3. The number of measurements that corresponds to the normal distribution is calculated for each interval. To accomplish this, the range of data is first centered and standardized.

Let $x_{\min} = a_0$ and $x_{\max} = b_0$, and divide the range $[a_0, b_0]$ into L intervals of length $h_0 = (b_0 - a_0)/L$. Centering and standardization are then achieved with the formula

$$x_{ic} = \frac{x_{i0} - \bar{x}}{\tilde{\sigma}}.$$

For example, the transformed limits of the range of the data for us will be as follows:

$$a_c = \frac{a_0 - \bar{x}}{\tilde{\sigma}}, \quad b_c = \frac{b_0 - \bar{x}}{\tilde{\sigma}}.$$

The length of the transformed interval $h_c = (b_c - a_c)/L$. Then we mark the limits $\{z_i\}$, $i = 0, 1, \ldots, L$, of all intervals of the transformed range $[a_c, b_c]$:

$$z_0 = a_c, \qquad z_1 = a_c + h_c, \qquad z_2 = a_c + 2hc, \cdots, z_L = a_c + Lh_c = b_c.$$

Now we calculate the probability that a normally distributed random quantity falls within each interval:

$$p_i = \frac{1}{\sqrt{2\pi}} \int\limits_{z_i}^{z_i+1} e^{-x^2/2} dx.$$

After this we calculate the number of measurements that would fall within each interval if the population of measurements is normally distributed:

$$\varphi_i = P_i n.$$

4. If less than five measurements fall within some interval, then this interval in both histograms is combined with the neighboring interval. Then the degree of freedom $\nu = L - 3$, where L is the total number of intervals (if the intervals are enlarged, then L is the number of intervals after the enlargement), is determined.
5. The indicator χ^2 of the difference of frequencies is calculated:

$$\chi_i^2 = \frac{(\tilde{\varphi}_i - \varphi_i)^2}{\varphi_i}, \qquad \chi^2 = \sum_{i=1}^{L} \chi_i^2.$$

6. The significance level of the test q is chosen. The significance level must be sufficiently small so that the probability of rejecting the correct hypothesis (committing false rejection) would be small. On the other hand, too small a value of q increases the probability of accepting the incorrect hypothesis, that is, of committing false retention.

From the significance level q and a degree of freedom ν in Table A.4, we find the critical threshold χ_q^2, so that $P\{\chi^2 > \chi_q^2\} = q$. The probability that the value obtained for χ^2 in step 5 above exceeds χ_q^2 is equal to q and is small. For this reason, if it turns out that $\chi^2 > \chi_q^2$, then the hypothesis that the distribution is normal is rejected. If $\chi^2 < \chi_q^2$, then the hypothesis that the distribution is normal is accepted.

The smaller the value of q, the larger is the value of χ_q^2 for the same value of ν, hence the more easily the condition $\chi^2 < \chi_q^2$ is satisfied and the hypothesis being tested is accepted. But, in this case, the probability of committing false retention increases. For this reason, q should not be taken to be less than 0.01. For too large a value of q, as pointed out above, the probability of false rejection increases and, in addition, the sensitivity of the test decreases. For example, for $q = 0.5$ the value of χ^2 may be greater or less than χ_q^2 with equal probability, and therefore it is impossible to accept or reject the hypothesis.

To achieve a uniform solution of the problem at hand, it would be desirable to standardize the significance levels q adopted in metrology.

It should be noted that the test examined above makes it possible to check the conformance of the empirical data to any theoretical distribution, not only a normal distribution. This test, however, as also, by the way, other goodness-of-fit tests, does not make it possible to establish the form of the distribution of the observations; it only makes it possible to check whether the observations conform to a normal or some other previously selected distribution.

3.7 Testing for Homogeneity of Samples

Measurements with large random errors require careful attention. One must make sure that the obtained results are statistically under control, stable, i.e., that the measurement results cluster around the same central value and have the same variance. If the measurement method and the object of investigation have been little studied, then the measurements must be repeated until one is sure that the results are stable [25]. This process determines the duration of the investigation and the required number of measurements.

The stability of measurements is often estimated intuitively based on prolonged observations. Mathematical methods exist that are useful for assessing the stability of measurements, so-called methods for testing homogeneity. A necessary condition for measurement stability is that the data passes the homogeneity tests. However, this is not sufficient for homogeneity in reality, because of a possibility of an unfortunate choice of groups of measurements.

Figure 3.3 shows the results of measurements of some quantities, presented in the sequence in which they were obtained. Consider three groups of measurements performed in the time intervals $t_2 - t_1$, $t_3 - t_2$, and $t_4 - t_3$. They apparently will be homogeneous. Meanwhile, subsequent measurements would differ significantly from the first measurements. On the whole, the results obtained from the first group of measurements will give a picture of a stable, statistically under control, measurement, which is actually not the case.

The choice of groups for monitoring homogeneity remains a problem for the experimenter. In general, it is best to have on the order of ten measurements in a group, and it is better to have several such groups than two groups with a large number of measurements. Once the groups have been reliably determined to be homogeneous, they can be combined and later regarded as one group of data.

We shall consider first the most common methods for testing homogeneity that assume the normal distribution of a population. These methods are called parametric; before using these methods, each group of data must first be checked for normality.

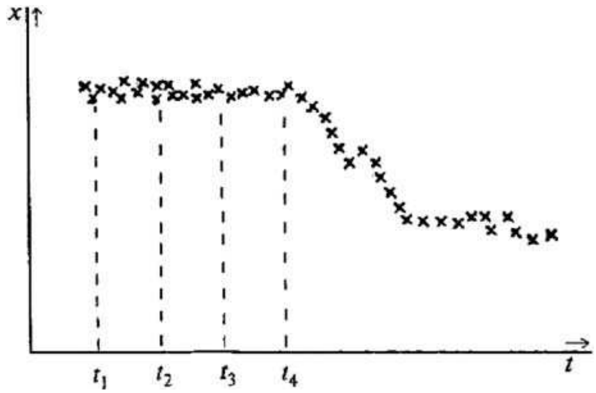


Fig. 3.3 Example of a sequence of single-measurement results obtained in an unstable measurement

The admissibility of differences between estimates of the variances is checked with the help of *Fisher's test* in the case of two groups of observations and *Bartlett's test* if there are more than two groups. We shall present both methods.

Consider two groups of observations, and let the unbiased estimates of the variances of these groups be S_1^2 and S_2^2, where $S_1^2 > S_2^2$. The number of observations in the groups is n_1 and n_2, so that the degrees of freedom for these groups are, respectively, $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. We form the ratio

$$F = \frac{S_1^2}{S_2^2}.$$

Next, from Tables A.5 and A.6, which present the probabilities $P\{F > F_q\} = q$ for different degrees of freedom ν_1 and ν_2 and for two values of q (1% and 5%), we choose the value F_q for a chosen value of q. The hypothesis is accepted, i.e., estimates of the variances can be regarded as corresponding to the same variance, if $F < F_q$. The significance level of the test, i.e., the probability of the wrong decision, is equal to $2q$.

Now assume that there are L groups. Assume unbiased estimates of the variances of groups of observations are known, $S_1^2, \ldots, S_L^2 (L > 2)$, and each group j has $\nu_j = n_j - 1$ degrees of freedom; in addition, all $\nu_j > 3$. The test of the hypothesis, that the variances of the groups are equal, is based on the statistic

$$M = N \ln \left(\frac{1}{N} \sum_{j=1}^{L} \nu_j S_j^2 \right) - \sum_{j=1}^{L} \nu_j \ln S_j^2,$$

where

$$N = \sum\nolimits_{j=1}^{L} \nu_j.$$

If the hypothesis that the variances are equal is correct, then the ratio

$$\chi_1^2 = \frac{M}{1 + \dfrac{1}{3\,(L-1)}\left(\sum\limits_{j=1}^{L} \dfrac{1}{\nu_j} - \dfrac{1}{N}\right)}$$

is distributed approximately as χ^2 with $\nu = L - 1$ degrees of freedom.

Given the chosen significance level q, from Table A.4, we find χ_q^2, such that $P\left\{\chi^2 > \chi_q^2\right\} = q$. If the inequality $\chi_1^2 < \chi_q^2$ is satisfied, then differences between the estimates of the variances are admissible, i.e., they could be due to randomness of the data.

The admissibility of differences between the arithmetic means is also checked differently in the case of two or more groups of observations. We shall first examine the comparison of the arithmetic means for two groups of observations, when there are many observations, so that each estimate of the variances can be assumed to be equal to its variance.

We denote by $\bar{x}_1$, σ_1^2, and n_1 the parameters of one group and by $\bar{x}_2$, σ_2^2, and n_2 the parameters the other group. We form the difference $\bar{x}_1 - \bar{x}_2$ and estimate its variance:

$$\sigma^2(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Next, having chosen a certain significance level q, we find $\alpha = 1–q$, and from Table A.1, we find the quantile $z_{\frac{1+\alpha}{2}}$ of the Gaussian function corresponding to the probability $\frac{1+\alpha}{2}$. A difference between the arithmetic means is considered admissible if

$$|\bar{x}_1 - \bar{x}_2| \le z_{\frac{1+\alpha}{2}}\sigma(\bar{x}_1 - \bar{x}_2).$$

If the variances of the groups are unknown (e.g., if the number of observations is not sufficient to take variance estimations for the values of variances), then the problem can be solved only if both groups have the same variances (the estimates of this variance $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ can, naturally, be different). In this case, the statistic

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(n_1 - 1)\tilde{\sigma}_1^2 + (n_2 - 1)\tilde{\sigma}_2^2}}\sqrt{\frac{n_1 n_2(n_1 + n_2 - 2)}{n_1 + n_2}}$$

is distributed approximately according to Student's distribution.

Then, given the significance level q, from Table A.2 for Student's distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom, we find t_q such that $P\{t > t_q\} = q$. The difference between the arithmetic means is regarded as admissible if $t < t_q$.

If the number of groups is large, the admissibility of differences between the arithmetic means is checked with the help of another variant of Fisher's test. The first step in Fisher's test includes a check that all groups have the same variance, using the methods above. Then, Fisher's method involves comparing estimates of the intergroup variance S_L^2 and the average variance of the groups $\bar{S}^2$:

$$S_L^2 = \frac{1}{L-1} \sum_{j=1}^{L} n_j \left(\bar{x}_j - \bar{x}\right),$$

where

$$\bar{x} = \frac{\sum\limits_{j=1}^{L} n_j \bar{x}_j}{N}, \quad N = \sum_{j=1}^{L} n_j$$

and

$$\bar{S}^2 = \frac{1}{N-L} \sum_{j=1}^{L} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right).$$

Both estimates of the variances have a χ^2 distribution with $\nu_1 = L - 1$ and $\nu_2 = N - L$ degrees of freedom, respectively. Their ratio has Fisher's distribution with the same degrees of freedom.

The spread of the arithmetic means is admissible if $F = S_L^2 \big/ \bar{S}^2$ for the selected probability α lies within the interval from F_L to F_U:

$$P\{F_L \leq F \leq F_U\} = \alpha$$

The upper limits of Fisher's distribution F_U are presented in Tables A.5 and A.6; the lower limits are found from the relation $F_L = 1/F_U$. If the significance levels in finding F_U and F_L are taken to be the same $q_1 = q_2 = q$, then the overall significance level of the test will be $2q$ and

$$\alpha = 1 - 2q.$$

A method for checking the admissibility of the spread in the arithmetic means of the groups when the variances of the groups are different has also been developed, but it is more complicated.

It should be noted that a significant difference between the arithmetic means could indicate that systematic errors exist in the observational results of some of the groups, and these errors are different in different groups. Therefore, measurements cannot be performed with the required accuracy.

We shall now discuss nonparametric methods for testing homogeneity. These methods do not require any assumptions about the distribution function of the population and are widely used in mathematical statistics.

We begin with *Wilcoxon rank sum test* for checking if two groups of observation belong to the same probability distribution. More formally, assume that we have two samples: $\{x_i\}$, $i = 1, \ldots, n_1$, of random quantity X, and $\{y_j\}$, $j = 1, \ldots, n_2$, of random quantity Y, and let $n_1 \leq n_2$. We check the hypothesis H_0: $F_1 = \mathrm{F}_2$, where F_1 and F_2 are the distribution functions of the random quantities X and Y, respectively.

The sequence of steps in checking H_0 is as follows. Both samples are combined, and an ordered series is constructed from $N = n_1 + n_2$ elements; i.e., all observations x_i and y_j are arranged in increasing order, irrespective of the sample to which these observations belong. Next, each element is assigned a *rank* as follows. Elements with unique values receive the rank equal to their order number in the series. All elements sharing the same values (which will obviously always appear next to each other in the series) receive the same rank equal to the arithmetic mean of their position numbers.

For example, the series (2.3, 2.5, 2.5, 2.6, 2.6, 2.6) will have ranks (1, 2.5, 2.5, 5,5,5). Indeed, the first element has a unique value so it receives its order number as its rank. The next two elements are equal and they get rank 2.5 equal to their average of their order numbers (2 and 3) in the series. The last three elements are also equal and receive the rank 5, which is the average of their positions (4, 5, and 6).

Next the sum of the ranks of all elements of sample $\{x_i\}$ is calculated. The sum T obtained is then compared with the critical value T_q for a selected significance level q. For small values of n_1 and n_2, tables listing $T_q(n_1, n_2)$ are given in most modern books on statistics. (These tables usually list values of T_q only for $n_1 \leq n_2$, which is why we compute T for the smaller sample.) For $n_1, n_2 > 25$, the critical value T_q can be calculated using the normal distribution $N(m_1,\ \sigma^2)$:

$$T_q = m + z_{1-q}\sigma,$$

where

$$m = \frac{n_1(N+1)}{2},\ \sigma^2 = \frac{n_1 n_2\,(N+1)}{12}$$

and z_{1-q} is the quantile of the standard normal distribution $N(0, 1)$ for probability $(1 - q)$. The hypothesis H_0 is rejected with significance level q against the alternative and it means that X is stochastically greater (i.e., has greater mathematical expectation) than Y if $T > T_q$. For a two-sided alternative, H_0 is rejected against the alternative that X is stochastically different from Y with significance level $2q$ if $T > T_q$ or if $T < n_1(N+1) - T_q$.

Another nonparametric method for checking homogeneity is *the Siegel–Tukey test*, which also considers two samples, $\{x_i\}$ and $\{y_j\}$, where $n_1 \leq n_2$ and tests the hypothesis H_0: $F_1 = \mathrm{F}_2$. The Siegel–Tukey test assumes that both distributions have the same mathematical expectation. All $N = n_1 + n_2$ values of the two samples are again arranged into one sequence in the increasing order, and each element is assigned a rank based on its position in the sequence. However, the procedure for rank assignment is different. First, preliminary ranks are assigned as follows: rank 1 is given to the first element, rank 2 to the last (Nth) element, rank 3 to the $(N-1)$st element, rank 4 to the second element, rank 5 to the third element, rank 6 to the $(N-2)$nd element, and so on. Then, all neighboring elements with equal values receive the same final rank equal to the average of the preliminary ranks of all these elements.

Next, we compute the sum R of the ranks of the elements of sample $\{x_i\}$. Assume for simplicity that samples are sufficiently large ($n_1, n_2 > 25$). From R, we calculate the standardized variable z, defined as

$$z = \frac{\left| R - \dfrac{n_1 (N+1)}{2} \right|}{\sqrt{\dfrac{n_1 n_2 (N+1)}{12}}}$$

For significance level q, the hypothesis H_0 is rejected if $z > z_{1-q}$, where z_{1-q} is a quantile for probability $(1-q)$ of the standard normal distribution $N(0,1)$.

The Wilcoxon's test is based on comparing the average values of two samples, whereas the Siegel–Tukey test is based on estimates of the variances. Indeed, in Wilcoxon's test, if the two expectations were dissimilar, observations of one sample would tend to group toward one side of the combined sequence. Then its rank sum T would tend to be either large or small. In contrast, ranks in Siegel–Tukey test are assigned so that elements away from the middle of the sequence receive smaller ranks than those close to the middle. If one sample had lower variance, its elements would tend to be clustered around the middle of the sequence. Thus, the sum of their ranks R would be high.

For this reason, these two tests supplement one another.

As an example of the complimentary nature of these tests, consider again the experiment from Sect. 2.7 that checked the homogeneity of two batches of the same types of measuring instruments. Table 3.1 gives calculation data for homogeneity checking of two batches of 160 ammeters for a moving-iron instrument Э59 with respect to the error at marker 30 of the graduated scale [47].

For the Wilcoxon's test, we obtain $T = 25,403$. Let $q = 0.05$. Then $z_{0.95} = 1.96$, and

$$T_q = \frac{160 \times 321}{2} + 1.96 \sqrt{\frac{160 \times 160 \times 321}{12}} = 27{,}620$$

As 25,403 $<$ 27,620, the hypothesis that the samples are homogeneous is accepted based on Wilcoxon's test.

Consider now the Siegel–Tukey test. According to the data in the table, $R = 23{,}713$. We thus obtain

$$z = \frac{\left| 23{,}713 - \dfrac{160 \times 321}{2} \right|}{\sqrt{\dfrac{160 \times 160 \times 321}{12}}} = 2.3.$$

Let us take $q = 0.05$ and therefore $z_{0.95} = 1.96$, the same values we used in the Wilcoxon's test. As $z > z_{0.95}$, the hypothesis that the samples are homogeneous is rejected based on the Siegel–Tukey test. Thus, the two tests bring different outcomes.

Table 3.1 The example of rank determination for nonparametric homogeneity testing

Value of the error	Number of instruments with a given error in the sample			Wilcoxon's test		Siegel–Tukey test	
	x	y	$x+y$	Average rank of a given value of the error	Sum of ranks for a given value of the error in the sample x	Average rank of a given value of the error	Sum of ranks for a given value of the error in the sample x
−0.50	1	1	2	1.5	1.5	2.5	2.5
−0.40	3	0	3	4.0	12.0	7.3	22.0
−0.30	3	0	3	7.0	21.0	13.7	41.0
−0.25	1	0	1	9.0	9.0	17.0	17.0
−0.20	13	5	18	18.5	240.5	36.5	474.5
−0.15	2	2	4	29.5	59.0	58.5	117.0
−0.10	10	8	18	40.5	405.0	80.5	805.0
−0.05	3	2	5	52.0	156.0	103.6	310.8
0.00	15	28	43	76.0	1,140.0	151.5	2,272.5
0.05	5	5	10	102.5	512.5	204.5	1,022.5
0.10	26	35	61	138.0	3,588.0	573.5	7,108.4
0.15	7	4	11	174.0	1,218.0	293.5	2,054.5
0.20	34	41	75	217.0	7,378.0	207.5	7,055.0
0.25	1	3	4	256.5	256.5	128.5	128.5
0.30	17	11	28	272.5	4,632.5	96.5	1,640.5
0.40	13	11	24	298.5	3,880.5	44.5	578.5
0.45	1	1	2	311.5	311.5	18.5	18.5
0.50	4	2	6	315.5	1,262.0	10.5	42.0
0.60	0	1	1	319.0	0.0	3.0	0.0
0.80	1	0	1	320.0	320.0	2.0	2.0

3.8 Robust Estimates

The distribution function by its nature is a mathematical concept. It is used in measurements as a theoretical model for a set of measurements. As always, a complete conformance between the model and the real set of data is impossible. Therefore, different models can be chosen for the same data. A small difference between the models may lead to significantly different estimation of the measurand. A solution to this problem was offered by so-called *robust estimations* [29, 32]. Among the earliest known robust estimations, the most popular are the truncated means, the Winsor's means, and the weighted means [32]. These methods assume that measurement results are arranged in an ordered series; i.e.,

$$x_1 \leq x_2 \leq \cdots \leq x_n.$$

- *The Truncated Means*. Given the ordered series above, the method of truncated means discards k values from the left and the right ends of this series. The number k is obtained as $k = \lfloor np \rfloor$, where $0 < p < 0.5$ and the notation $\lfloor np \rfloor$ means that k is the greatest integer number that is equal to or smaller than np. The rest of the series provides the robust estimate of the measurand by the formula

$$\tilde{A}_T = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_i .$$

Note that the truncating procedure is similar to the usual practice of eliminating the outlying result from the sample, which is described in Sect. 3.4.

- *The Winsor's Means.* Rather than discarding extreme items in the ordered series, the Winsor's method replaces them with the neighboring items. The robust estimate of the measurand is calculated by the formula:

$$\tilde{A}_W = \frac{1}{n} \left\{ \sum_{i=k+1}^{n-(k+1)} x_i + (k+1)(x_{k+1} + x_{n-k}) \right\} .$$

- *The Weighted Means.* The weighted means method obtains a robust estimate by computing a linear combination of the measurement data. There are numerous variations in this method [29,32]. Here we present one such variation, which uses the weighted average of the median of the series and two items symmetrically located around the median in the series [32].

Median M is determined by the formula:

$$M = \begin{cases} x_{k+1} & \text{if } n = 2k+1; \\ \frac{1}{2}(x_k + x_{k+1}) & \text{if } n = 2k. \end{cases}$$

The robust estimate of the mean according to this method is then given by the following formula:

$$\tilde{A}_C = (1-2\epsilon)M + 2\epsilon \frac{(x_l + x_{n-l+1})}{2},$$

where $(1-2\varepsilon)$ and 2ε are the weights, $\varepsilon \ll 1$, and l and $(n-l+1)$ are the positions of the two symmetrical items chosen for the estimation.

Numerous other robust estimates were also proposed. Thus, it is not clear which method to choose for a given measurement. Hogg [29] addressed this difficulty as follows. His method takes advantage of the natural assumption that all density distributions are symmetrical, the assumption on which all other robust estimates are based anyway. Symmetrical distributions can be characterized by one parameter – the excess e (see Sect. 3.1):

$$e = \frac{\mu_4}{\sigma^4}.$$

Hogg proposed to divide all distributions into several classes depending on the value of e, in such a way that for all distributions in the same class, the mean value can be calculated with the same formula. Thus, the estimate of the measurand for each class will not depend on the distribution function. The estimate of the excess e is found from the formula:

$$æ = \frac{\sum_{i=1}^{n}(x_i - \tilde{A})^4}{nS^4}.$$

The price this method pays for the robust estimate is the loss in the efficiency of the estimate. Therefore, a desired solution would find a compromise between the number of classes and the loss of the efficiency. Hogg studies the system of four classes named classes *A, B, C,* and D. The range of values of æ for each class and the corresponding formulas for estimating the mean value of the data are given in Table 3.2. Hogg found that the four classes he proposed lead to loss in efficiency of no more then 20%, which is acceptable.

Another system of classes was proposed later by Mechanikov [39]. This system contains only three classes, which are also determined by the values of æ. These classes and the corresponding formulas for the estimation of the mean are shown in Table 3.3. As one can see, the formulas in Table 3.3 are the same as those used in the Hogg system: Class 1 uses the same formula as Class D, Class 2 as Class B, and Class 3 as Class A, but Class C is eliminated.

The estimations of variances of robust estimates are calculated in a common way, but constructing confidence intervals presents a difficult problem that is generally not discussed in the robust estimates literature. A simple nonparametric (i.e., not relying on a particular probability distribution) method to construct these intervals has been proposed in [28]. In this method, the confidence interval is defined by two elements located symmetrically about the median in the ordered series.

Table 3.2 Classes of distribution functions and formulas for estimation of their mean values after Hogg

Distribution class	æ	Formula for the measurand estimation
A	$æ < 2$	$\tilde{A}_a = \frac{1}{2}(x_1 + x_n)$
B	$2 < æ < 4$	$\tilde{A}_b = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
C	$4 < æ < 5.5$	$\tilde{A}_c = \frac{1}{n-2\lfloor n/4\rfloor}\sum_{i=\lfloor n/4\rfloor+1}^{n-\lfloor n/4\rfloor} x_i$
D	$5.5 < æ$	$\tilde{A}_d = M$

Table 3.3 Classes of distribution functions and formulas for estimation of their average values after Mechanikov

Distribution class	æ	Formula for the measurand estimation
1	$4 < æ$	$\tilde{A}_{1m} = M$
2	$2.5 < æ < 4$	$\tilde{A}_{2m} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
3	$1.8 < æ < 2.5$	$\tilde{A}_{3m} = \frac{x_1 + x_n}{2}$

For a given confidence probability α, the symmetrical positions l and r, which define the confidence interval $[x_l, x_r]$, are found as follows:

$$l = \lfloor \frac{1}{2}(n+1 - z_{\frac{1+\alpha}{2}}\sqrt{n}) \rfloor \text{ and } r = \left\lceil \frac{1}{2}\left(n + 1 + z_{\frac{1+\alpha}{2}}\sqrt{n}\right)\right\rceil,^2$$

where $z_{\frac{1+\alpha}{2}}$ is the corresponding quantile of the standard normal distribution.

For example, for the ordered series of size $n = 49$ and $\alpha = 0.95$, $A \approx M = x_{25}$ and $l = 19$ and $r = 31$. The confidence interval is thus $[x_{19}, x_{31}]$.

The inverse calculation was proposed in [39]. Here, we first choose the symmetrical elements in the ordered series as the confidence interval boundaries and then calculate the corresponding confidence probability for this interval. Let k be the distance of the boundary elements from their corresponding ends of the sequence, so that the interval is $[x_k, x_{n-k+1}]$. The confidence probability that the true value A is covered by that confidence interval is computed according the formula:

$$P\{x_k \le A \le x_{n-k+1}\} = \frac{1}{2^n} \sum_{i=k}^{n-k+1} \binom{i}{n}.$$

In particular, for

$$k = 2, \qquad P\{x_2 < A < x_{n-1}\} = 1 - \frac{n+1}{2^{n-1}},$$

$$k = 3, \qquad P\{x_3 < A < x_{n-2}\} = 1 - \frac{n^2+n+2}{2^n}.$$

For $k > 3$, the formulas become much more complicated. But for $k = 4$ and 5, one can use approximate relations presented in [39]:

$$k = 4, \qquad P\{x_4 < A < x_{n-3}\} \approx 1 - \frac{0.17n^3}{2^{n-1}},$$

$$k = 5, \qquad P\{x_5 < A < x_{n-4}\} \approx 1 - \frac{0.037n^4}{2^{n-1}}.$$

Nonparametric methods are widely used in statistical analysis. However, to construct confidence intervals, they require many more observations than parametric methods.

Another way to build confidence intervals is made possible by a *bootstrap method* [23]. This method uses a computer to produce a large number of independent replicas of the obtained experimental dataset (e.g., 1,000 replicas). For each replica, we compute an estimate of the measurand as the arithmetic mean for the replica. We produce so many of these estimates that they can be assumed to represent well the distribution function of the estimates. Then, we can use this distribution function

---

[2] As usual, $\lfloor x \rfloor$ denotes the greatest integer equal to or smaller than x and $\lceil x \rceil$ stands for the smallest integer equal to or greater than x.

(which we can obtain, e.g., by guessing it from the estimates and then checking our guess using statistical methods for hypothesis testing) to compute the final estimate of the measurand and its confidence interval with usual methods.

3.9 Application of the Bayes' Theorem

The Bayes' Theorem is well studied in the probability theory. Also widely held among mathematicians has been an opinion that this theorem allows one to utilize a priori information about the measurand and in this way to improve the accuracy of the measurement. Further, it is appealing to consider a measurement as a process of increasing the amount of acquired information and, correspondingly, of increasing the accuracy of the obtained results.

The initial or a priori information in the Bayes' Theorem is usually considered to be the probability density function of the measured quantity [21]. Unfortunately, this information is not, and cannot be, available. Perhaps for this reason the Bayes' Theorem did not find practical usage in measurement data processing until recently.

A real possibility to use Bayes' Theorem in metrological practice was opened by research based on the concept of *likelihood* [31, 36]. Following the monograph [36], the propositions of interest in metrological applications are usually (a) the measurand Q belongs to an infinitesimal interval $(q, q + \mathrm{d}q)$ and (b) d is the data obtained in the result of the measurement.

Let $f(q)$ be the PDF of measurand Q before the measurement; it represents a priori knowledge about Q, and $f(q|d)$ be the conditional PDF of Q given the measurement data d. Then, according to Bayes' Theorem,

$$f\,(q|d) = \frac{f(q)f(d|q)}{f(d)}.$$

Integrating both parts of the above equation by q, under the assumption that $f(d)$ is constant, and after applying the normalization condition that $\int\limits_{-\infty}^{+\infty} f\,(q|d)\,\mathrm{d}q = 1$, we can obtain

$$f(d) = \int\limits_{-\infty}^{+\infty} f\,(q|d)\, f(q)\mathrm{d}q.$$

It is suggested to consider $f(d|q)$ as the PDF of variable d assuming Q takes given values q, if d is referred to the possible values of some random quantity D. To return to the original meaning of notations d and q, a function l is introduced. Function l differs from f in that its arguments d and q switch places; furthermore, l is defined so that

$$f(d|q) = l(q|d).$$

Function l is called *likelihood*. With its introduction, Bayes' Theorem takes the form

$$f(q|d) = \frac{l(q|d)\,f(q)}{\int\limits_{-\infty}^{+\infty} f\,(q|d)\,f(q)\mathrm{d}q}.$$

Monograph [36] points out that the new function cannot be considered as a PDF but it represents a new concept, which is called likelihood. This concept is then applied to a direct multiple measurement and several indirect measurements.

Let us consider the direct measurement. Its a priori information is that the measurement method employed produces observations that belong to a normal distribution. The monograph compares the results obtained using the modified Bayes' Theorem with the results produced by a traditional method of maximum likelihood with the same normal distribution of the observations.

It turned out that while both methods produce the same estimate of the measurand, their estimates of the variance are different. The estimate produces using the modified Bayes' Theorem is

$$S^2\,(\bar{q}) = \sqrt{(n-1)/(n-3)} * S^2,$$

where n is the number of repeated measurements in the multiple measurement, S^2 is the variance estimate produced by the maximum likelihood method and $S^2\,(\bar{q})$ is the same estimate produced by the new method based on the modified Bayes' Theorem.

The increase in the variance estimate is small but significant, and this discrepancy requires an explanation. First, it is noteworthy that while the primary motivation for using the Bayes' Theorem was to extract more accuracy from the measurement data, the variance estimate it produced turned out to be higher, meaning the opposite outcome. Moreover, long practice of utilizing the maximum likelihood method has not given reason to suspect that it produces results with artificially overestimated accuracy. Second, both methods cannot be correct given that they produce different variance estimates. These issues must be resolved before one can recommend applying the Bayes' Theorem in practical measurements.

Chapter 4
Direct Measurements

4.1 Relation Between Single and Multiple Measurements

The classical theory of measurement errors is constructed based on the well-developed statistical methods and pertains to multiple measurements (we refer the reader back to Chap. 1 for the introduction of basic terms such as multiple and single measurements, uncertainty, error, and limits of errors). In practice, however, the overwhelming majority of measurements are single measurements, and however strange it may seem, for this class of measurements, there is no accepted method for estimating their inaccuracy.

In searching for a solid method for estimating errors in single measurements, it is first necessary to establish the relation between single and multiple measurements. At first glance, it seems natural to regard single measurements as a particular case of multiple measurements, when the number of measurements is equal to 1. Formally this is correct, but it does not serve any purpose, because statistical methods do not work for single observations. In addition, the question of when one measurement is sufficient remains open. In the seemingly natural approach above, to answer this question – and this is a fundamental question – it is first necessary to perform a multiple measurement and then, analyzing the results, to decide whether a single measurement was possible. But such an answer is in general meaningless: A multiple measurement has already been performed, and nothing is gained by knowing, in the hindsight, one measurement would have sufficed. Admittedly, it can be countered that such an analysis will make it possible not to make multiple measurements when future such measurements are performed. Indeed, that is how the above approach is used, but only when preliminary measurements are performed, i.e., in scientific investigations when some new object is studied. This is not done in practical measurements.

When one needs to measure, for example, the voltage of some source with a given accuracy, they choose a voltmeter with suitable accuracy and perform the measurement. If, however, the numbers on the voltmeter indicator dance about, then it is impossible to perform a measurement with the prescribed accuracy, and one must reexamine the measurement task and objective rather than performing a multiple measurement.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_4,

For practical applications, we can state the opinion that single measurements are well grounded in experience, distilled in the construction of the corresponding measuring instruments, and measuring instruments are manufactured so that single measurements could be performed.

From the foregoing assertion, a completely different point of view follows regarding the relationship between single and multiple measurements. Namely, single measurements are the primary, basic form of measurement, whereas multiple measurements are derived from single measurements, and in essence, they are simply repeated single measurements. Multiple measurements are performed when necessary, based on the formulation of the measurement problem. It is interesting that measurement problems that require multiple measurements are known beforehand; they can even be enumerated. Namely, multiple measurements are performed in the following cases:

1. When investigating a new phenomenon or a new object and relationships between the quantities characterizing the object, as well as their connection with other quantities, are being determined; in other words, when preliminary measurements, according to the classification given in Chap. 1, are performed.
2. When measuring the average value of some parameter, according to the goal of the measurement problem.
3. When the effect of random errors of measuring instruments must be reduced.

There is another point of view, namely, that any measurement must be a multiple measurement, because otherwise it is impossible to judge the measurement process and its stability and to estimate its inaccuracy. We cannot agree with this opinion. First, it contradicts practice, where single measurements dominate. Second, it also does not withstand fundamental analysis.

Imagine that the same constant quantity is measured simultaneously using a multiple and a single measurement. In both cases, the measurements are performed with the same analog instrument whose response time is t_r. In Fig. 4.1a, the dots show the

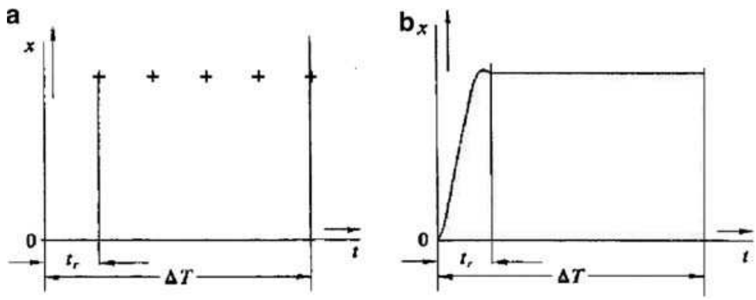


Fig. 4.1 Results of measurements in the case of (**a**) a multiple measurement and in (**b**) a single measurement with continuous photorecording of the indication

results of individual measurements comprising the multiple measurement, and the curve in Fig. 4.1b represents a continuous photorecording of the indications of the instrument in the single measurement. The single measurement makes it possible to obtain the value of the measurand immediately after the instrument response time t_r, while the multiple measurement takes at least this time multiplied by the number of individual measurements.

If it is desirable to check the stability of the measurement, then one can continue the observation using the single measurement. The measurement process is stable if the readings of the instrument over a chosen time ΔT do not change appreciably.

Furthermore, it is possible to estimate the inaccuracy of the result of a single measurement. Methods for calculating errors and uncertainty of the results of single measurements are given later in this chapter. Thus, in this case, a single measurement is sufficient to obtain the measurement result, to estimate its inaccuracy, and to assess the stability of the measurement process. In fact, a single measurement allows one to make a better judgment than a multiple measurement because the latter represents only separate moments of the process, whereas the former gives the whole continuous picture.

The above example does not say that a single measurement is better than a multiple measurement. It says only that a multiple measurement should not be performed when a single measurement is possible. But when a multiple measurement is necessary, a single measurement cannot possibly replace it, and in this case and in this sense, a multiple measurement is better than a single measurement.

Yet the above example supports our argument that single measurements must be regarded as independent and the basic form of measurement. Correspondingly, the problem of developing methods for estimating the accuracy of single measurements must be regarded as an independent and important problem of the theory of measurements.

This is a good point at which to discuss another aspect of the question at hand. In many fields of measurements, modern digital measuring instruments can operate so fast that over the time allotted for a measurement, say, 1 s, hundreds of measurements can be performed. By carrying out these measurements and averaging their results, we utilize all of the time allotted for measurement, and, thanks to this, we reduce correspondingly the effect of interference and noise.

Consider now an analog instrument having the same accuracy as a fast measuring device, but with the response time equal to the time allotted to the measurement, i.e., in our case, 1 s. From the time constant of the instrument, the effect of interference and noise will be suppressed to the same degree as for discrete averaging in the first case; i.e., we shall obtain the same result.

In other words, the measurement time is of fundamental importance, and there is no significance in how the interference and noise are filtered – in the discrete or analog form – over this time. In practice, discrete averaging is often more convenient, because in this case, the averaging time can be easily changed.

4.2 Classification of Elementary Errors

The classification of measurement errors presented in Chap. 1 also applies, of course, to elementary errors. Continuing the analysis, this classification must be further developed as it applies to elementary errors. The main two types of elementary errors are systematic and random errors.

Taking into account and eliminating systematic errors is an important problem in every accurate measurement. In the theory of errors, however, little attention has been devoted to systematic errors. In most books on methods of data processing, the question of systematic errors is either neglected or it is assumed that these errors have been eliminated. In reality, however, systematic errors cannot be completely eliminated; some unexcluded residuals always remain. These residuals must be taken into account to estimate the limits of the unexcluded systematic error of the result.

In addition, many measurements are performed without special actions taken to eliminate systematic errors, because either it is known a priori that they are small or the measurement conditions make them impossible to be eliminated. For example, in measurements of the mass of a body, corrections are often not made for the values of the balance weights employed, either because the corrections are small or because the errors of the weight values are unknown (only their limits are known).

Sometimes the unexcluded residuals of the systematic errors are assumed to be random errors based on the fact that their values are unknown. We cannot agree with this point of view. When classifying errors as systematic or random, attention should be focused on their properties rather than on whether their values are known.

For example, suppose that the resistance of a resistor is being measured and a correction is made for the influence of the temperature. The systematic error would be eliminated if we knew exactly the temperature coefficient of the resistor and the temperature. But we only know both quantities with limited accuracy, and for this reason, we cannot completely eliminate this error. An unexcluded residual of the error will remain. It can be small or large; this we can and should estimate, but its real value remains unknown. Nonetheless, this residual error has a definite value, which remains the same when the measurement is repeated under the same conditions, and for this reason, it is a systematic error.

Errors that have been eliminated are no longer errors. Therefore, the unexcluded residuals become the systematic error in the measurement if they cannot be neglected.

The error in a measurement can be both systematic and random, but after the measurement has already been performed, the measurement error becomes a systematic error. Indeed, the result of a measurement has a definite numerical value, and its difference from the true value of the measured quantity is also constant. Even if the entire error in a measurement was random, for a measurement result, it becomes systematic; i.e., it seemingly freezes.

We shall now discuss the classification of systematic errors. Our discussion on systematic errors classification is based on the work of M.F. Malikov, and following this work, we shall distinguish systematic errors according to their sources and properties [37].

The sources of systematic errors can be three components of the measurement: the method of measurement, the measuring instrument, and the experimenter. Correspondingly, methodological, instrumental, and personal systematic errors are customarily distinguished.

Methodological errors arise from imperfections of the method of measurement and from the limited accuracy of the formulas used to describe the phenomena on which the measurement is based. We shall also classify as methodological errors the errors arising as a result of the influence of the measuring instrument on the object whose property is being measured.

For example, the moving-coil voltmeter draws current from the measurement circuit. Because of the voltage drop on the internal resistance of the source of the voltage being measured, the voltage on the terminals of the voltmeter will be less than the measured value. The indications of the voltmeter, however, reflect the voltage on its terminals. The error that arises – a methodological error – should be insignificant or eliminated by a correction.

A methodological error can also arise in connection with the use of the measuring instrument. For example, the gain of a voltage amplifier is determined by measuring the voltages at the input and the output. If these voltages are measured successively using the same voltmeter, as is often done in practice, then, aside from the voltmeter error, the measurement error will include the error from some uncontrollable change in voltage at the amplifier input over time. This error does not arise when two voltmeters are employed to measure the input and output voltage at the same time. (Of course, in the case of the two voltmeters, the overall measurement error is impacted by the instrumental errors of both of the voltmeters, so the choice of the measurement method must depend on the particular circumstances. For instance, if the input voltage was known to be stable, the one-voltmeter method would be preferable.)

We note that the error from the threshold discrepancy between the model and the object (see Sect. 1.4) is also a methodological error.

Instrumental systematic errors are errors caused by imperfections of the measuring instrument. One example of such errors is errors caused by imprecise calibration of the instrument scale. Other examples include the inaccuracy of balance weights and the error of a resistive voltage divider from the inaccurate adjustment of the resistances of its resistors.

Another group of such errors is additional and dynamic errors. These errors also depend on the imperfections of the measuring instruments, but they are caused by influence quantities and noninformative parameters of the input signal (see Sect. 2.3) as well as by the change in the input signal in time. Most often the additional and dynamic errors are systematic errors. When the influence quantities and the forms of the input signal are unstable, however, they can become random errors.

Setup errors, i.e., errors arising from the arrangement of the measuring instruments in conducting the measurement and their effect on one another, are also instrumental errors.

Personal systematic errors are systematic errors caused by the individual characteristics of the observer. Specifically, we shall discuss the errors in the reading of the indications of indicating instruments. Such errors were investigated by H. Bäkström [17]. He studied the question of how people estimate tenths of the graduations of an instrument scale when reading the instrument indication. Although Bäkström's work simulated real devices by drawings depicting the edges of a scale graduation and the indicator of the instrument, the results obtained are plausible.

In his study, Bäkström presented the drawings to human subjects and asked them to estimate the tenths of the graduation given by the indication. He found that the systematic errors made by every observer when estimating tenths of a graduation of an instrument scale can reach 0.1 of the graduation and are much larger than random errors. These systematic errors are manifested by the fact that for different positions of the indicator within the graduation, different observers characteristically produce estimates with different frequencies, and in addition, the distribution characteristic of the estimates for every observer remains constant for a long period of time. This phenomenon can be explained by the conjecture that one observer tends to refer indications relative to the lines forming the edges of graduation and to the middle (fraction 0.5) of a graduation. Another observer refers indications to the fractions 0.4 and 0.6 of a graduation. A third observer prefers fractions 0.2 and 0.8 of graduations and so on.

The error in estimation of tenths of graduations depends on the thickness of the markers – the lines forming the scale. The optimal thickness of these markers is 0.1 of the length of a graduation. The length of a graduation also significantly affects the error in reading tenths of a graduation. Instrument scales for which tenths of a graduation can be read are usually made so that the length of a graduation is equal to about 1 mm (not less than 0.7 mm and not more than 1.2 mm). On the whole, for a random observer, the distribution of systematic errors in the readings of tenths of a graduation can be assumed to be uniform with limits of ± 0.1 graduations.

Let us now consider types of systematic errors according to their properties. In this regard, constant systematic errors are distinguished from regularly varying systematic errors. The latter, in turn, are subdivided into progressing and periodic errors and errors that vary according to a complicated law.

A *constant systematic error* is an error that remains constant, and for this reason, it is repeated in each observation or measurement. For example, such an error will be present in measurements performed using the same instruments and devices that have a systematic error: balance weights, measuring resistors, and so on. The personal errors made by experienced experimenters can also be classified as constant (for inexperienced experimenters, they are usually of a random character).

Progressing errors are errors that increase or decrease with passing of time, so every later observation will have a higher or lower error. Such errors are caused, for example, by the change in the working current of a potentiometer from the voltage drop of the storage battery powering it.

Periodic errors are errors that vary with a definite period. In the general case, a systematic error can vary according to a complicated aperiodic law.

The detection of systematic errors in a measurement is a complicated problem. It is especially difficult to detect a constant systematic error. To solve this problem, several measurements (at least two) should be performed by fundamentally different methods. This method is ultimately decisive. It is often realized by comparing the results of measurements of the same quantity that were obtained by different experimenters in different laboratories.

It is easier to discover variable systematic errors, which can be done with the help of statistical methods, correlation, and regression analysis. But nonmathematical possibilities also should not be avoided. Thus, in the process of performing a measurement, it is helpful to employ a graph on which the results of the measurements are plotted in the sequence in which they were obtained. The overall arrangement of the points obtained makes it possible to discover the presence of a systematic change in the results of observations without mathematical analysis. If a regular change in observational results has been found and it is known that the measured quantity did not change in the process, then this indicates the presence of a regularly varying systematic error. The human capability of perceiving such regularities is widely employed in metrology, although this capability has apparently still not been thoroughly studied.

It is also helpful to measure the same quantity using two different instruments (methods) or to measure periodically a known quantity instead of the unknown quantity.

If the presence of a systematic error has been discovered, then it can usually be estimated and eliminated. In precise measurements, however, this often presents great difficulties and is not always possible.

In most fields of measurements, the most important sources of systematic errors are known and measurement methods have been developed that eliminate the appearance of such errors or prevent them from affecting the result of a measurement. In other words, systematic errors are eliminated not by mathematical analysis of experimental data but rather by the use of appropriate measurement methods. The analysis of measurement methods and the systematization and generalization of measurement methods are important problems, but they fall outside the scope of this book, which is devoted to the problem of analysis of experimental data. For this reason, we shall confine our attention to a brief review of the most widely disseminated general methods for studying such problems.

Most constant systematic errors are estimated analytically before the measurement and not from the experimental data obtained during the measurement. These a priori estimates usually produce definite (nonprobabilistic) limits for these errors. We shall further divide constant systematic errors into absolutely constant and conditionally constant errors.

By *absolutely constant elementary errors*, we mean errors that, although they are specified by definite limits, remain the same in repeated measurements performed under the same conditions as well as when using different measuring instruments of the same type. Consider for example a thermocouple. The errors of thermocouples of each type are rated by specifying their standard characteristic (the dependency of the output EMF on the temperature difference at input). Every point of this characteristic has its own error, which is constant for this point. There are known limits of error for the thermocouple characteristic as a whole, so that the error at any point of the characteristic falls within these limits. This information should be taken into account when estimating the inaccuracy of the measurement of temperature.

By *conditionally constant errors*, we mean errors that have definite limits but can vary within these limits due to the individual properties of particular measuring instruments used in the measurement. A typical example of such an error is the measurement error caused by the intrinsic error of the measuring instrument.

The intrinsic error, by its nature, can be a purely systematic error, but it can also have a random component. For example, for weights, the intrinsic error does not have a random component, but the actual magnitude of the intrinsic error varies from one weight to another. The intrinsic error of an electric measuring instrument with an indicator needle has both systematic and random components, but on the whole, the intrinsic error has definite limits that are the same for any instrument of a given type.

A conditionally constant error can even be purely random. Examples are the rounding error in reading the indications of analog instruments and the error caused by the limited resolution of digital instruments.

In summary, a fundamental property of conditionally constant elementary errors is that although they have definite limits, they can vary within these limits.

Let us now turn to random errors. Before we proceed, it is interesting to note that the random errors are usually not classified into categories based on their causes, because a random error occurs in the course of a multiple measurement and is not predicted from an a priori analysis like systematic errors.

The random error is estimated using data obtained in the course of the measurements. If the random error is significant, then the measurement is performed many times. The primary characteristic of a random error is usually the standard deviation, which is calculated from the experimental data. The entire standard deviation, and not its separate components, is estimated directly. For this reason, there is no need to qualify the term *random measurement error* with the additional word *elementary*.

When performing an analysis, it is important to distinguish purely random and quasirandom errors. Purely random errors can arise from different reasons. For example, they can arise from noise or small (regarded as permissible) variations in the influence quantities or the random components of the errors of the measuring equipment.

Quasirandom errors appear in measurements of quantities that are by definition averages, when the quantities being averaged are constant. As the simplest (albeit artificial) example, one could measure a side of a (assumed to be) square object as the average of its all four sides. Each side will be somewhat different from the others, but will remain constant.

With quasirandom errors, the differences between individual quantities being averaged are not random but are regarded as random. Using this assumption, the quasirandom error of the measurement result can be characterized, just as in the case of a purely random error, by an estimate of the standard deviation.

4.3 Modeling of Elementary Errors

Ultimately, elementary errors are needed to assess the overall inaccuracy of the measurement, which usually means estimating the uncertainty of the measurement result. In other words, the measurement uncertainty is calculated from the elementary errors that are components of the overall measurement inaccuracy; i.e., this is a problem of synthesis, performed mathematically. Correspondingly, elementary errors must be represented by mathematical models. We shall examine the most common types of elementary errors (according to their properties) from this viewpoint: absolutely constant errors, conditionally constant errors, purely random errors, and quasirandom errors. We will not consider models of the variable, progressing, and periodic systematic errors because it is impossible to specify general models for these types. Thus, these errors should be taken into account differently in each particular case.

4.3.1 Absolutely Constant Errors

Each such error has a constant value that is the same in any measurement, although it is unknown. Only the limits of these errors are known. But since this error is constant within the known limits, the probabilistic model is not suitable in this case. A mathematical model of such errors should rather be considered a determinate quantity whose magnitude has an interval estimate; i.e., it lies within an interval of known limits. We shall use this model for absolutely constant elementary errors.

We can foresee an objection to this model. There is an opinion that if the value of the error is unknown, then it can be regarded as a random quantity. However, this is not correct. A model of an object can be constructed only based on what we know about it and not based on what we do not know.

There is another objection. If the determinate model above is adopted, then when several absolutely constant errors are summed, their limits must be added arithmetically. This process is equivalent to the assumption that all terms have limiting values and the same sign, which is unlikely. The objection then is that the determinate model leads to overestimation of the overall measurement inaccuracy. This objection also is invalid. First, the argument "unlikely" is not correct here, because we are not using a probabilistic model. Second, the fact that we do not like the result – the answer seems exaggerated – is also not an argument. In mathematics, precisely the same situation arises in methods of approximate calculations and the limits of errors are added arithmetically.

Fortunately, in a measurement, rarely more than one or two absolutely constant errors exist, and they are, as a rule, insignificant. Thus, summing their limits arithmetically does not usually lead to overly exaggerated uncertainty in practice.

4.3.2 Conditionally Constant Errors

The values of these errors characteristically vary from one measurement to another and from one measuring instrument to another, and they are different under different conditions. In all cases, however, in each such error, the limits of the interval containing any possible realization of such an error remain unchanged.

As a mathematical model of conditionally constant errors, one would like to use a random quantity. To specify this model, however, it is necessary to know the probability distribution function of this random quantity. Ideally, one would like to find this function based on the experimental data. Such an attempt was made for the intrinsic error of measuring instruments. The results of such an investigation were presented in Chap. 2. Unfortunately, they showed that the distribution function of the intrinsic error and, of course, the distribution function of the additional errors could not be found from sample data.

Thus, to adopt the probabilistic model for conditionally constant errors, the distribution function must be prescribed. It is well known that among distributions with fixed limits, the uniform distribution has the highest uncertainty (in the sense of information theory). As an analogy, the rounding error also has known limits, and in mathematics, this error has for a long time been regarded as a random quantity with a uniform probability distribution. For this reason, we shall also assume that the model of conditionally constant errors will be a random quantity with a uniform probability distribution within prescribed limits.

This suggestion was made a long time ago [48]. At the present time, this model is widely employed in the theory of measurement errors [2, 5, 11].

4.3.3 Purely Random Errors

Such errors, often referred to as just "random errors" for short, appear in multiple measurements. They are characterized by the standard deviation that is computed from the experimental data.

The form of the distribution function of random errors can, in principle, be found based on the data from each multiple measurement. In practice, however, the number of measurements performed in each experiment is insufficient for this. Thus, every time measurements are performed, it is assumed that the purely random errors have a normal distribution, relying on the implicit assumption is that the hypothesis of the normal distribution was checked in a preceding experiment. Unfortunately, the normal distribution hypothesis is rarely directly checked. Yet the results obtained using these assumptions are not inconsistent with the practice so that this assumption is evidently justified. Thus, we shall assume that the mathematical model of random errors is, as a rule, a *normally distributed random quantity.*

4.3.4 Quasirandom Errors

As noted above, these errors occur when measuring quantities that are averages by definition, and the value of each separate quantity being averaged remains constant. These quantities are essentially not random, but can sometimes be regarded as a random sample from a general population of quantities. Whether or not such an assumption is justified depends on the goal of the measurement, and it is a judgment call based on agreement of experts. If one does assume the randomness of the underlying quantities, the parameters to be used to characterize their distribution should also be determined by agreement. Most often the standard deviation is chosen as this parameter.

We will conclude this section with a discussion on the question of interdependence and correlation of elementary errors. Mathematically, it is preferable to regard these errors as correlated quantities, because this approach is extremely general. However, such an approach complicates the inaccuracy estimation, and most of the time it is not justified. Under reference conditions, all elementary errors are independent and thus are uncorrelated. Exceptions can be encountered in measurements performed under rated operating conditions, especially in the case of indirect measurements and measurements performed with the help of measuring systems, when the same influence quantity causes appreciable additional errors in several instruments or components in the measuring channel of the system. An example is a measurement in which a measuring transducer, amplifier, and automatic-plotting instrument are employed. A change in the temperature of the medium can cause these devices to acquire an additional temperature-induced error. Obviously, these additional errors will be interrelated. Accounting for the dependency between additional errors is considered in Chap. 5.

4.4 Composition of Uniform Distributions

In Sect. 4.3, we have adopted the uniform distribution as the mathematical model of conditionally constant elementary errors. Given several conditionally constant elementary errors that contribute to the overall measurement error, how can we assess the overall error? As already mentioned, this is a problem of synthesis of the overall error from its components. To solve this problem, one must know how to construct the composition of uniform distributions. The theoretical solution of this problem is well known and is presented, for example, in [53]. However, our applied problem at hand allows us to construct a simplified solution. We will consider this solution in the current section, and then, in subsequent sections, use the described apparatus to estimate the inaccuracy of direct measurements.

Consider n random quantities $x_i (i = 1, \ldots, n)$, each of which has a uniform distribution centered at zero in the interval $\left[-\frac{1}{2}, +\frac{1}{2}\right]$, and denote $\vartheta = \sum_{i=1}^{n} x_i$. The probability density function of the sum of these random quantities has the form

$$f_n(\vartheta) = \frac{1}{(n-1)!}\left[\left(\vartheta + \frac{n}{2}\right)^{n-1} - \binom{n}{1}\left(\vartheta + \frac{n}{2} - 1\right)^{n-1} + \binom{n}{2}\left(\vartheta + \frac{n}{2} - 2\right)^{n-1} - \cdots\right]$$

where the sum must include only the terms in which power bases, i.e., $\vartheta + \frac{n}{2}$, $\vartheta + \frac{n}{2} - 1$, and so on, are nonnegative. Note that the number of terms therefore depends on both the number of components being summed, n, and the argument ϑ. For example, if $n = 2$, then

$$f_2(\vartheta) = (\vartheta + 1) - 2\vartheta = \begin{cases} 0, & \vartheta \leq -1, \\ \vartheta + 1, & -1 < \vartheta \leq 0, \\ 1 - \vartheta, & 0 \leq \vartheta < 1, \\ 0, & 1 < \vartheta. \end{cases}$$

The probability density function of the sum of two terms has the form of a triangle. For $n = 3$, the graph of $f_3(\vartheta)$ consists of three segments of a quadratic parabola and looks very much like the curve of a normal distribution. For $n = 4$, this distribution is almost indistinguishable from the normal distribution. Given the above equation for the probability density, it is not difficult to find the probability distribution function

$$F_n(\vartheta) = \frac{1}{n!}\left[\left(\vartheta + \frac{n}{2}\right)^n - \binom{n}{1}\left(\vartheta + \frac{n}{2} - 1\right)^n + \binom{n}{2}\left(\vartheta + \frac{n}{2} - 2\right)^n - \cdots\right] \tag{4.1}$$

In practice, however, it is desirable to have a simpler and more convenient solution. Such a solution can be found by observing that we only need to find the confidence interval for the combined error and not its full distribution function. In other words, we are interested in limits $\pm\theta_\alpha$ for the sum of the components such that the probability

$$P\{|\vartheta| \leq \theta_\alpha\} \geq \alpha.$$

Bearing this in mind, we shall examine the distribution function $F_n(\vartheta)$ in the extreme intervals of its argument range with nonzero probability density, $[-n/2, -n/2 + 1]$ and $[n/2 - 1, n/2]$.

For these intervals, (4.1) assumes the form

$$F_n(\vartheta) = \begin{cases} \frac{1}{n!}\left(\vartheta + \frac{n}{2}\right)^n & \text{for } -\frac{n}{2} < \vartheta < -\frac{n}{2} + 1, \\ 1 - \frac{1}{n!}\left(\vartheta - \frac{n}{2}\right)^n & \text{for } \frac{n}{2} - 1 < \vartheta < \frac{n}{2}. \end{cases}$$

The composition of the distributions is symmetric relative to the ordinate axis. We shall discuss how to calculate, given the probability distribution, the limits of the confidence interval corresponding to a fixed value α of the confidence probability. The limits of the confidence interval corresponding to α are $\pm\theta_\alpha$.

By definition, the probability that the true value of a quantity ϑ lies within the confidence interval $[-\theta_\alpha, +\theta_\alpha]$ is α. Therefore, the probability that the quantity does not lie in the confidence interval is $(1 - \alpha)$. If the distribution is symmetric relative to 0 (and we are studying a symmetric distribution), then the probability that the quantity will take on a value less than $-\theta_\alpha$ will be equal to the probability that it will take on a value greater than $+\theta_\alpha$. These probabilities are obviously equal to $(1-\alpha)/2$.

Consider first the left-hand branch of the distribution function. The probability corresponding to the point $-\theta_\alpha$ is equal to $P\{\vartheta \leq -\theta_\alpha\} = (1-\alpha)/2$. Considering now the right-hand branch, the probability that $\vartheta \leq +\theta_\alpha$ will obviously be equal to $1 - [(1-\alpha)/2] = (1+\alpha)/2$.

We shall now return to our problem. Given $F_n(\vartheta)$ and α, we are required to find the quantiles $-\theta_\alpha$ and $+\theta_\alpha$ (recall that the quantile of a distribution function for a given probability level is the argument on which the distribution function takes the value equal to the specified probability level). Since these quantiles have equal absolute values, we shall only calculate $-\theta_\alpha$.

Since the confidence probability is usually high (e.g., 0.95), quantile $-\theta_\alpha$ is likely to fall into the left extreme interval $[-n/2, -n/2+1]$ (we can check if that is indeed the case once we calculate it, or even beforehand as we will see shortly). Then, we have

$$P\{\vartheta \leq \theta_\alpha\} = F_n(-\theta_\alpha) = \frac{1}{n!}\left(-\theta_\alpha + \frac{n}{2}\right)^n = \frac{1-\alpha}{2}, \tag{4.2}$$

from which θ_α can be calculated.

For example, let $\alpha = 0.99$ and $n = 4$. Then $(1-\alpha)/2 = 0.005$. Let us check whether the value $(-\theta_\alpha)$ corresponding to this probability falls within the left extreme interval $[-2, -1]$. To do so, we can simply find the value of the cumulative distribution function for the upper limit of this interval, i.e., -1:

$$F_4(-1) = \frac{1}{4!}(-1+2)^4 = \frac{1}{1 \times 2 \times 3 \times 4} = 0.041.$$

As $0.005 < 0.041$, and because the cumulative distribution function is a monotonically growing function, we know that the value $(-\theta_\alpha)$ is less than (-1) and hence lies in the interval $[-2, -1]$.

We shall represent θ_α found from formula (4.2) in the following form:

$$\theta_\alpha = k\sqrt{\sum_{i=1}^{n} \theta_i^2}, \tag{4.3}$$

where θ_i represents the range of each component error x_i, $(-\theta_i \leq x_i \leq +\theta_i)$, and k is a correction factor. In the case at hand, $\theta_i = 1/2$ for all $i = 1, \ldots, n$; i.e.,

$$\theta_\alpha = k\frac{\sqrt{n}}{2}, \quad k = 2\theta_\alpha/\sqrt{n}. \tag{4.4}$$

Formula (4.3) is convenient for calculations, and for this reason, we shall investigate the dependence of the coefficient k on α and n. The calculations are performed as follows. Given α and n, we find θ_α from (4.2). Next, the correction factor k is found for the given values of α and n from formula (4.3).

Table 4.1 Values of the coefficient k for various number of component errors and confidence probability

Number of component errors, n	Values of the coefficient k for confidence probability α			
	0.90	0.95	0.99	0.9973
2	0.97	1.10	1.27	1.34
3	0.96	1.12	1.37	1.50
4	*	1.12	1.41	1.58
5	*	*	*	1.64
...	...	...	...	...
∞	0.95	1.13	1.49	1.73

*These values are not calculated because critical values $-\vartheta_\alpha$ and $+\vartheta_\alpha$ fall outside the through extreme intervals of the cumulative distribution function domain

Continuing with our example of $\alpha = 0.99$ and $n = 4$, we find θ_α by substituting these values into (4.2):

$$\frac{1}{4!}\left(-\theta_\alpha + 2\right)^4 = 0.005, \quad \theta_\alpha = 2 - \sqrt[4]{24 \times 0.005} = 1.41.$$

Having found θ_α, we obtain from formula (4.4):

$$k = \frac{2 \times 1.41}{\sqrt{4}} = 1.41.$$

Table 4.1 presents the values of k for other values of α and n; these values were calculated similarly to the method above. The value of k for $n \to \infty$ was found using the fact that by the central limit theorem, the resulting distribution can be assumed normal.

Recalling the notation $\vartheta = \sum_{i=1}^{N} x_i$, we can obtain the standard deviation of ϑ as follows:

$$V[\vartheta] = V\left[\sum_{i=1}^{n} x_i\right] = \sum_{i=1}^{n} V[x_i].$$

But, as is well known, $V[x_i] = \theta_i^2/3$. Therefore

$$V[\vartheta] = \frac{\sum\limits_{i=1}^{i} \theta_i^2}{3}, \qquad \sigma[\vartheta] = \sqrt{\frac{1}{3}\sum_{i=1}^{n} \theta_i^2}. \tag{4.5}$$

Furthermore, the mathematical expectation of ϑ is zero because the mathematical expectation of each x_i is zero. Thus, if $n \to \infty$, we have random quantity ϑ with a normal distribution $N(0, \sigma)$. We can then now calculate the absolute value of the limits of the confidence interval as $\theta_\alpha = z_\text{p}\sigma$, where z_p is the quantile of the standard normal distribution $N(0, 1)$ corresponding to the probability $p = (1 + \alpha)/2$ (see above for the explanation of computing probability p). Thus, we obtain

$$\theta_\alpha = \frac{z_p}{\sqrt{3}} \sqrt{\sum_{i=1}^{n} \theta_i^2} \tag{4.6}$$

Comparing (4.6) with (4.3), we find

$$\underset{n\to\infty}{k} = \frac{z_\mathrm{p}}{\sqrt{3}}.$$

For example, for $\alpha = 0.9973$, we obtain $z_\mathrm{p} = 3$ and thus, when the number of component errors is large, $k = 1.73$.

Considering Table 4.1, one can observe that the correction factor k has the interesting property that for $\alpha \leq 0.99$, it is virtually independent of the number of components. We can make use of this property and take for k the average values in each column. These values of k are given in Table 4.2.

The error caused by using the average values of k, as one can see by comparing them with the exact values given in Table 4.1, does not exceed 10% for $\alpha = 0.99$ and 3% for $\alpha = 0.95$.

The small effect of the number of components indicates indirectly that it is not always necessary to assume, as was done above, that all θ_i are equal. For instance, assume that one of the limits, θ_l, is gradually reduced. The effect on factor k will be negligible because even in the extreme, when θ_l is reduced all the way to zero and the lth component disappears, the values of k for $(n - 1)$ and n components are virtually the same. If, on the other hand, θ_l is gradually increased, then the factor k will decrease.

Figure 4.2 depicts the dependence of k on the ratio $c = \theta_l/\theta_0$ for $\alpha = 0.99$, where θ_0 is the absolute value of the remaining terms, which are assumed to be equal. This figure can be used to find k more precisely than using Table 4.2. The figure also shows that for every n, coefficient k is at the maximum when all θ_i are equal.

Factor k can also be calculated using formulas approximating the curves presented in Fig. 4.2. For $\alpha = 0.99$ and $n = 4$, a good approximation formula is

$$k = 1.45 - 0.05\frac{\theta_l}{\theta_0}.$$

Formula (4.6) can be used instead of (4.3) to calculate θ_α when the number of terms is large. However, as follows from the above-presented estimate of the error of calculations based on formula (4.3), the accuracy cannot be increased by more than

Table 4.2 Average values for coefficient k

α	0.90	0.95	0.99
k	0.95	1.10	1.40

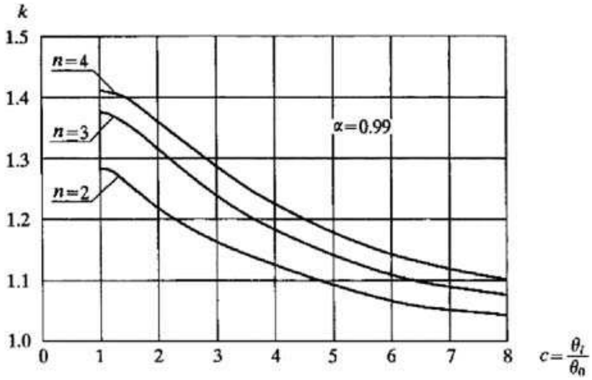


Fig. 4.2 Coefficient k as a function of the change in limits of one of the component errors relative to the other component errors (the number of components $n = 2, 3, 4$)

10% (for $\alpha = 0.99$). At the same time, formula (4.3) is also useful for summing a small number of terms. For this reason, for practical calculations, relation (4.3) is preferable.

With a confidence probability $\alpha = 0.99$ and $n \leq 4$, it could turn out that our approximate calculation of θ_α would produce $\theta_\alpha > \sum_{i=1}^{n} \theta_i$. But this obviously cannot happen. In this case, one can take

$$\theta_\alpha = \sum\nolimits_{i=1}^{n} \theta_i.$$

Of course, a more correct alternative in using the above value would be to obtain a more accurate value of the coefficient k from the curves in Fig. 4.2.

There arises, however, the question of how well founded the confidence probability choice $\alpha = 0.99$ is. In most cases, this limit does not correspond to the reliability of the initial data, and the limit $\alpha = 0.95$ is more appropriate. For $\alpha = 0.95$, Table 4.2 gives $k = 1.1$, and formula (4.3) assumes the form

$$\theta_{0.95}V = 1.1\sqrt{\sum_{i=1}^{n} \theta_i^2}.$$

In this case, $\theta_\alpha < \sum_{i=1}^{n} \theta_i$ always holds. To see this, first let $n = 2$ and assume without loss of generality that $\theta_1 \leq \theta_2$. It is not difficult to verify that the inequality $\theta_\alpha = 1.1\sqrt{\theta_1^2 + \theta_{21}^2} < (\theta_1 + \theta_2)$ holds as long as $\theta_1/\theta_2 > 0.11$. But the last condition is always satisfied in practice because an elementary error that is about ten times smaller than any other elementary error can be neglected.

Consider now three components, and assume $\theta_3 \geq \theta_2 \geq \theta_1$. Denoting $T = \theta_3 + \theta_2$, we obtain an equivalent inequality

$$1.1\sqrt{T^2 + \theta_1^2 - 2\theta_3\theta_2} < (T + \theta_1).$$

The term $2\,\theta_3\theta_2 > 0$, and therefore it is enough to prove the above inequality without this term under the square root (indeed, if the simplified inequality holds, the original inequality will only be stronger). Then, similar to the case with two components we have just studied, we can show that the simplified (and hence the original) inequality holds as long as

$$\frac{\theta_1}{\theta_2 + \theta_3} > 0.11.$$

It is obvious that this condition holds easier than for two components and is always satisfied in practice. On the whole, as the number of component elementary errors increases, the inequality $\theta_\alpha < \sum_{i=1}^{n} \theta_i$ is satisfied only more easily. Since we showed that this inequality is satisfied in practice even for two components, we can conclude that it always holds in practice for an arbitrary number of components.

4.5 Methods for Precise Measurements

Methods for precise measurements attempt to eliminate systematic errors. They also reduce random errors by means of repeating the measurement many times and statistical processing of the obtained results. The most common methods for precise measurements are the following.

Method of replacement. This method involves replacing the quantity to be measured with a known quantity in a manner so that no changes occur in the indication of all measuring instruments employed. Then, we can assume that the measured quantity is equal to the known quantity that replaced it. The method of replacement is the most accurate method of measurement.

Consider, for example, *Borda's method* for weighing. The method is designed to eliminate the systematic error from the inequality of the arms of the balance. Let x be the measured mass, P be the mass of the balancing weights, and l_1 and l_2 be the lengths of the arms of the balances. The measurement is performed as follows. First, the body being weighed is placed in one pan of the balance and is balanced with the help of a weight with mass T. Then,

$$x = \frac{l_2}{l_1} T.$$

Next, the mass x is removed and a known mass P that once again balances the pans is placed in the empty pan:

$$P = \frac{l_2}{l_1} T.$$

As the right-hand sides of both equations are the same, the left sides are also equal to one another, i.e., $x = P$, and the fact that $l_1 \neq l_2$ has no effect on the result.

The resistance of a resistor can be measured in an analogous manner with the help of a sensitive but inaccurate bridge and an accurate magazine of resistances. Several other quantities can be measured analogously.

Method of contraposition. The measurement is performed with two observations, and it is performed so that the reason for the constant error would affect the results of observations differently but in a known, regular fashion.

An example of this method is *Gauss's method* of weighing. First, the body being weighed is balanced by balance weights P_1. Using the notation of the preceding example, we have

$$x = \frac{l_2}{l_1} P_1.$$

Next the unknown weight is placed into the pan that previously held the balancing weights and is again balanced by the balance weights. Now we have

$$x = \frac{l_2}{l_1} P_2.$$

We now eliminate the ratio l_2/l_1 from these two equalities and find

$$x = \sqrt{P_1 P_2}.$$

The sign method of error compensation. This method involves two measurements performed so that the constant systematic error would appear with different signs in each measurement.

For example, consider the measurement of electromotive force (EMF) x with the help of a DC potentiometer that has external wires with a parasitic thermo-EMF. One measurement gives E_1. Next, the polarity of the measured EMF is reversed, the direction of the current in the potentiometer is also reversed, and once again the measured EMF is balanced. This process gives E_2. If the thermo-EMF produces error ϑ and $E_1 = x + \vartheta$, then $E_2 = x - \vartheta$. From here,

$$x = \frac{E_1 + E_2}{2}.$$

Elimination of progressing systematic errors. The simplest and most frequent case of a progressing error is an error that changes linearly in proportion to time. An example of such an error is the error in the measurement of voltage with a potentiometer, if the voltage of the storage battery, generating the working current, drops appreciably.

Formally, if it is known that the working current of the potentiometer changes linearly in time, then to eliminate the arising error, it is sufficient to perform two observations at known times after the working current along the standard cell is regulated. Let

$$E_1 = x + Kt_1, \qquad E_2 = x + Kt_2,$$

where t_1 and t_2 are the time intervals between regulation of the working current and the observations, K is the coefficient of proportionality between the measurement error and the time, x is the voltage being measured, and E_1 and E_2 are the results of the observations. From the above equations, we obtain

$$x = \frac{E_1 t_2 - E_2 t_1}{t_2 - t_1}.$$

For accurate measurements, however, it is best to use a somewhat more complicated *method of symmetric observations*. In this method, several observations are performed equally separated in time and then the arithmetic means of the pairs of symmetric (i.e., the first and last, the second and the second-to-last, etc.) observations are calculated. Theoretically, with linearly changing systematic errors, these averages must be equal, which makes it possible to control the course of the experiment and to eliminate these errors.

4.6 Accuracy of Single Measurements Using Measuring Instruments Under Reference Conditions

The great majority of measuring instruments were created for single measurements. Some of these instruments are so simple that the inaccuracy of corresponding measurements can be estimated without calculation. For example, the inaccuracy of the length measurement performed with a ruler is determined simply by rounding the readings on the ruler. Also, calculating the inaccuracy is not necessary when it is known beforehand that the accuracy of that measurement will be "good enough" for the goal of this measurement. This includes most of the household measurements, such as measuring the voltage of a car battery with an industrial tester or weighing the ingredients for a cooking recipe. In other measurements, the inaccuracy must be calculated.

Under reference conditions, the inaccuracy of single measurement is determined by the limits of the intrinsic error: there are no additional errors by definition. The limits of the intrinsic errors of measuring instruments are known; they are listed in the documentation provided by the manufacturers or in the certificates from the calibration laboratories. The problem is only to recalculate these limits, if necessary, for a given indication of the instrument, i.e., for the measurement result.

If the limits of the intrinsic error are given in the form of absolute or relative errors and are the same for the whole range of the instrument, then recalculations are not required and these limits are the limits of the given elementary error. But often the limits of intrinsic error of a measuring instrument are given in the form of a fiducial error, i.e., as a percentage of the fiducial value. The conversion into relative error is then made using the formula

$$\delta_{\text{in}} = \gamma \frac{x_N}{x}. \tag{4.7}$$

where δ_{in} is the limit of the intrinsic error in relative form, γ is the limit of the fiducial error, x_N is the fiducial value, and x is the reading of the instrument in the corresponding units. Conversion into the form of absolute errors is done according to the formula

$$\Delta_{\text{in}} = \delta_{\text{in}} x = \gamma x_N. \tag{4.8}$$

It was mentioned in Sect. 2.3 that the fiducial errors are expressed in percents. Therefore, to obtain Δ_{in} in proper form of absolute errors, it must be divided by 100.

When the estimate of inaccuracy of a single measurement is obtained using the limits of intrinsic errors listed in the manufacturer's documentation, the estimate remains correct even if the instrument used in the measurement is replaced with another instrument of the same type. Indeed, the limits of the intrinsic error listed in the manufacturer's documentation apply to all instruments of this type. Recall that measurement inaccuracy estimates obtained from such data were termed universal in Chap. 1. In contrast, the estimates obtained using data from a certificate of a calibration laboratory that applies to a specific instrument were called individual.

In some cases, a measurement error may arise from the interaction between the object of study and the measuring instrument employed. For instance, when measuring an electric voltage with an indicating voltmeter, the voltmeter reacts on the strength of the electric current it consumes, and as it was mentioned above in Sect. 4.2, its indication shows not the voltage being measured but the voltage on the voltmeter's terminals. This creates a systematic error, which depends on the relative values of the input impedance of the voltmeter and the internal impedance of the source of the voltage being measured. Most often, this error is negligibly small. But in some cases it needs to be taken into account and be compensated with a correction. Then only the error of the correction will remain as a contributing factor in the inaccuracy of the measurement. We consider in detail an example of this kind of error in Sect. 8.1.

We shall now consider several examples of calculating the universal estimates of the inaccuracy of single measurements.

1. Industrial tester WV-531A (RCA). This is a multifunctional instrument, and its accuracy is different for different measurement ranges. Let us assume, for example, that we need to measure the AC voltage using the 150 V range. The manufacturer specification says that the instrument's inaccuracy in this range for AC voltage measurements is ±4% of the full-scale value.

   So, we have here the limits of fiducial error $\gamma = \pm 4\%$ and the fiducial value $x_N = 150$ V. Assume the instrument indication in our measurement was 117.5 V. In accordance with (4.7), the limits of error of this measurement result are

   $$\delta = \pm 4\% \times \frac{150}{117} = \pm 5\%.$$

   In the form of absolute error, these limits are

   $$\Delta = \frac{4\% \times 150}{100\%} = \pm 6\,\text{V}.$$

   Thus, the result of this measurement must be presented as

   $$118\,\text{V} \pm 5\% \text{ or } (118 \pm 6)\,\text{V}.$$

2. Fluke 5700 A [26]. Assume we need to perform a measurement at the scale range of 11 V. The limits of intrinsic error at this range are $\pm$(5 ppm of output $+4\,\mu$V).
   If the indication of the instrument in our measurement is 10.000463 V, then the limits of error of this measurement will be

$$\Delta = \pm(10.000463 \times 5 \times 10^{-6}\text{V} + 4\,\mu\text{V}) = \pm 54\,\mu\text{V}.$$

   Since this can be considered a precise measurement, we can retain both significant digits in the inaccuracy above and present the measurement result as (10.000463 ± 0.000054) V.
3. Consider the digital multirange voltmeter example from Chap. 2 with specifications listed again below:

Time after calibration	24 h	90 days	12 months
Temperature	$23 \pm 1°$C	$23 \pm 5°$C	$23 \pm 5°$C
10.00000 V	–	–	$\pm$(35ppm +5ppm)
1000.000 V	$\pm$(20ppm +6ppm)	$\pm$(35ppm +10ppm)	$\pm$(45ppm +10ppm)

We refer the reader to Chap. 2 for the clarifications on the meaning of the entries in this table. We will only recall here that when the error of an instrument is listed using two terms as in this table, the first term expresses the error relative to the instrument indication, while the second term, even though it is expressed in the relative form, is *not* a relative error. As explained in Chap. 2, this term is a fiducial error and is expressed relative to the value that corresponds to the end of the measurement range of the instrument; this error is therefore the same for any indication in the entire range even when recalculated to the absolute form.

Assume the voltmeter is used to measure 500.0 V immediately after calibration and then again 12 months later, both times under reference conditions. Using the above specification (in particular, the columns corresponding to 24 h and 12 months since calibration), we shall evaluate the limits of absolute measurement error in both cases. Note that since the instrument is used under reference condition, the last column of the specification is not considered.

For the first measurement, we have:

$$\Delta_1 = \pm(500 \times 20 \times 10^{-6} + 1000 \times 6 \times 10^{-6})\text{V} = \pm 16\,\text{mV}.$$

After 12 months, the limits of error become:

$$\Delta_2 = \pm(500 \times 45 \times 10^{-6} + 1000 \times 10 \times 10^{-6})\text{V} = \pm 32.5\,\text{mV}.$$

4.7 Accuracy of Single Measurements Using Measuring Instruments Under Rated Conditions

When measurement is performed under rated operating conditions, the measurement result, as before, is given by the instrument indication. However, the calculation of the measurement inaccuracy turns into a more complex problem. Solving this problem starts with estimating the elementary errors of the measurement.

It is difficult to formulate a single method for estimating elementary errors, because these errors are by their nature extremely diverse. The general recommendations for solving this problem can nonetheless be formulated.

To estimate elementary measurement errors, it is first necessary to determine their possible sources. If it is known that some corrections will be (or have been) introduced, then the errors in determining the corrections must be included among the elementary errors.

All elementary measurement errors must be estimated in the same manner, i.e., in the form of either absolute or relative errors. Relative errors are usually more convenient for a posteriori error estimation, and absolute errors are more convenient for a priori error estimation. However, the tradition of each field of measurement should be kept in mind. Thus, for lineal–angular measurement, absolute errors are typically used, whereas for measurements of electromagnetic quantities, relative errors are preferred.

An unavoidable elementary error in any measurement is the intrinsic error of the measuring instrument. We presented the methodology of recalculating the intrinsic error of the instrument into the elementary error of the measurement in Sect. 4.6.

Additionally, the environmental conditions, characterized by the temperature, pressure, humidity, vibrations, and so on, also affect the result of a measurement. Each influence quantity, in principle, engenders its elementary error. To estimate it, it is first necessary to estimate the possible value of the corresponding influence quantity and then compare it with the limits of the range of values of this quantity concerning the reference condition. If the influence quantity falls outside the limits of reference values, then it causes a corresponding additional error; this error is also an elementary error.

Consider an error due to the temperature. Let the temperature of the medium exceed its reference values by ΔT. If, according to the rated operating conditions, the limit of the additional error due to ΔT is the same for an interval $T_1 \leq \Delta T \leq T_2$, then this limit is the limit of the given additional error. If, however, for this interval, the upper bound of the temperature coefficient is given, then the limits of temperature error are calculated according to the formula

$$\delta_T = \pm \omega_T \Delta T,$$

where δ_T is the limit of additional temperature error in the relative form and w_T is the upper bound of the absolute value of the temperature coefficient of the instrument expressed as the percentage of the instrument indication.

In general, for influence quantity i, the dependence of the limit of additional error δ_i or Δ_i on the deviations of the influence quantity outside the limits of its reference values can be given in the form of a graph or expressed analytically. In either case, the manufacturer's specifications of the instrument sometimes provide the influence function in the form of two components – the nominal influence function and an admissible deviation from it. This form allows one to take into account the deviation from the reference range by the corresponding correction to the measurement result. In the process, the elementary error decreases significantly, even if the influence function is specified with a large margin of error.

Suppose, for example, instead of the upper bound of the temperature coefficient w_T, the temperature coefficient is listed in the form $w'_T = (1 \pm \varepsilon)\, w_{T,N}$, where $w_{T,N}$ is the nominal temperature coefficient and ε is the admissible deviation from it, expressed in the relative form as a fraction of $w_{T,N}$. For temperature deviation ΔT from the upper limit of reference range, T, the additional error will be

$$\delta_T = w_{T,N}\Delta T \pm \varepsilon w_{T,N}\Delta T. \tag{4.9}$$

Because the first term in the above equation reflects a deterministic nominal dependency, we can account for it with the help of the correction

$$c = -w_{T,N}\Delta T \times b,$$

where b is the instrument indication. There then remains the temperature error

$$\delta'_T = \pm\varepsilon w_{T,N}\Delta T. \tag{4.10}$$

Even if the influence function is listed comparatively inaccurately, for example $\varepsilon = 0.2$ (20%), the temperature error still decreases greatly, by a factor of 4–6 in this case:

$$\frac{\delta_T}{\delta'_T} = \frac{1 \pm 0.2}{0.2} = 4 \text{ or } 6.$$

Finally, one should keep in mind that if the influence quantity itself is estimated with an appreciable error, then this error must also be taken into account when calculating the corresponding additional error.

In many cases, the input signal in a measurement is a function of time and therefore the measurement result may have a dynamic error. This error is also an elementary error that needs to be taken into account. Unfortunately, although the treatment of dynamic elementary errors has been discussed in research literature (e.g., [27]), the proposed methods are not mature enough to include here.

Once the errors of a single measurement have been analyzed, we have an estimate of the limits of all elementary errors of the measurement. We now proceed to the problem of synthesis, that is, the calculation of the overall inaccuracy of the measurement. In general, this calculation can be done using the following step-by-step procedure.

1. Identify all possible sources of elementary errors. The list of elementary errors always includes the intrinsic error of instrument involved and additional errors

due to influence quantities whose values fall outside the limits of the reference condition. Also, the interaction between that instrument and the object whose parameter is being measured, the discrepancy between the object and its model, and so on, must be taken into consideration.

2. Estimate the limits of all elementary errors. General recommendations to accomplish this step were described earlier. If point estimates have been obtained for some elementary errors, then one must apply the corresponding corrections to the instrument indication. In this case, the inaccuracy of the corrections must be taken into account along with the other elementary errors. We gave an example of a correction and of accounting for its inaccuracy earlier in this section, when considering the nominal temperature coefficient of an instrument. Another example can be found in Sect. 8.1.
3. Express the estimates of all elementary errors in the same form, either absolute or relative. Note that, as discussed in Sect. 4.6, the intrinsic error is often expressed as fiducial error. In this case, the fiducial error must be recalculated to the absolute or relative error of the measuring instrument reading in the actual measurement in question.
4. Calculate the inaccuracy of the measurement result. The procedure for this calculation is described next.

When one comes to step 4, all elementary errors have been estimated with their limits. Further calculations will require us to distinguish conditionally constant errors, absolute constant errors, and random errors. In single measurements, the vast majority of elementary errors are conditionally constant errors. Random errors are usually insignificant and can be accounted for as part of those conditionally constant errors in which they manifest themselves. Absolute constant errors occur infrequently.

We will begin with the conditionally constant errors. Among them, let ζ_0 be the intrinsic error of the measuring instrument and ζ_i, $i = 1, \ldots, m$, be the other elementary errors.

We now need to combine, or "sum up" these errors:

$$\zeta = \zeta_0 + \sum_{i=1}^{m} \zeta_i, \tag{4.11}$$

where ζ is the overall conditionally constant error. We know the limits θ_0 and θ_i of the elementary errors:

$$|\zeta_0| \leq \theta_0 \text{ and } |\zeta_i| \leq \theta_i .$$

Combining the elementary errors is often done by summing up their limits arithmetically. This is obviously the safest estimate, reflecting the worst-case scenario that all conditionally constant errors simultaneously reached their upper or lower limits. However, unlike in the case of absolute constant errors (where the errors are what they are and thus the question about the practicality of a particular combination of error values is invalid), the above scenario is unacceptable in the case of conditionally constant errors. A more realistic solution to this problem is provided by a probabilistic approach. To this end, we can utilize the mathematical model that we accepted for conditionally constant errors, which is to consider them as random variables uniformly distributed within their limits. If we in addition assume that these random variables are independent,[1] we can apply the discussion from Sect. 4.4 to calculate the measurement uncertainty as follows.

According to Sect. 4.4, the measurement uncertainty can be calculated using simple formula (4.3), which in our case is more convenient to rewrite to form:

$$u_\alpha = k_\alpha \sqrt{\theta_0^2 + \sum_{i=1}^{m} \theta_i^2}. \tag{4.12}$$

The analysis of Sect. 4.4 showed that for the most common confidence probability $\alpha = 0.95$, coefficient $k_{0.95} = 1.1$ and, remarkably, its value is independent of the number of components $n = m+1$. The inaccuracy of using (4.12) with this constant value for k is less then 3%. For $\alpha = 0.99$, if we assume $k_{0.99} = 1.4$, the inaccuracy of the calculation using (4.12) ranges from +10% for $n = 2$ to -6% for n tending to infinity.

One can increase the accuracy of this calculation in the last case by utilizing Table 4.1 or the graph on Fig. 4.2 to select the specific value of coefficient k for the measurement at hand. However, when the number of component errors is five or higher, it is justified in practice (and more convenient) to follow the analysis from Sect. 4.4 for the case of a large number of variables, which assumes that the combined variable has a normal distribution.

According to (4.5), the variance σ^2 of the resulting error can be obtained as

$$\sigma^2 = \theta_0^2/3 + \frac{1}{3}\sum_{i=1}^{m} \theta_I^2. \tag{4.13}$$

Knowing the variance and the shape of the distribution function, one can construct the confidence interval that covers the true value of the measurand with a given confidence probability α, i.e., to calculate the uncertainty of the measurement result as follows:

$$u_\alpha = z_p \sigma, \tag{4.14}$$

where z_p is the quantile of the standard normal distribution for probability $p = \frac{(1+\alpha)}{2}$. For $\alpha = 0.95$, (4.14) brings a well-known result $u_{0.95} = 1.96$, and for $\alpha = 0.99$ $u_{0.99} = 2.58\sigma$.

We would like to conclude the discussion of combining conditionally constant errors with an important practical recommendation. As we mentioned in Sect. 4.4, when the number of component errors is particularly small, i.e., four or less, and $\alpha \geq 0.99$, it is possible that the probabilistically combined error could produce an exaggerated estimate, which can even exceed the arithmetic sum of the component errors. Thus, for small number of components, it is advisable to combine the

[1] This assumption in fact follows naturally from the way instrument's additional errors are rated separately for individual influence quantities. However, further discussion on the validity of this assumption is outside the scope of this book.

elementary errors in *both* ways, arithmetically and probabilistically, and use as the result the smaller of the two uncertainty values produced. Note that this does not contradict the principle of upper-bound error estimates because the error can never exceed the arithmetic sum of its components.

Now consider the case where the measurement also has an absolutely constant error, in addition to conditionally constant errors we just examined. As we already mentioned, absolutely constant errors are relatively rare. In any case, one instrument can introduce only one absolutely constant error component to the overall measurement inaccuracy. If the absolutely constant error has limits H, then the overall measurement uncertainty will be

$$U_t = H + u_\alpha.$$

Because absolutely constant errors are the same in all instruments of the same type, these errors cannot be described using a probabilistic model. Thus, we have no choice but to add the limits of these errors arithmetically to the probabilistic sum of the conditionally constant errors.

It could happen that m of the n conditionally constant errors have asymmetric limits:

$$\theta_{jl} \le \vartheta_j \le \theta_{jr},\ j = 1, \ldots, m,$$

where θ_{jl} is the left-hand limit and θ_{jr} is the right-hand limit of component error j. The remaining $(n - m)$ conditionally constant errors are symmetric:

$$-\theta_j \le \vartheta_j \le \theta_j, j = m + 1, \ldots, n.$$

For calculations, asymmetric limits must be represented as symmetric limits around center a_j, where

$$a_j = \frac{\theta_{jl} + \theta_{jr}}{2}.$$

The limits of the interval that is symmetric relative to a_j are calculated according to the formula

$$\theta_j = \frac{\theta_{jr} - \theta_{jl}}{2}.$$

Note that the above calculation cannot be used to transform asymmetric errors into symmetric by introducing corrections into the measurement results: The error estimates are too unreliable to change the measurement result.

Next, the limits of the overall conditionally constant error must be calculated from the following formulas:

$$\theta_{r,\alpha} = \sum_{j=1}^{m} a_j + k\sqrt{\sum_{j=1}^{m} \theta_j^2 + \sum_{j=m+1}^{n} \theta_j^2},$$

$$\theta_{l,\alpha} = \sum_{j=1}^{m} a_j - k\sqrt{\sum_{j=1}^{m} \theta_j^2 + \sum_{j=m+1}^{n} \theta_j^2}. \tag{4.15}$$

Table 4.3 A fragment of specification of a multirange voltmeter

Time after calibration	24 h	90 days	12 months	Temperature coefficient
Temperature	23 ± 1°C	23 ± 5°C	23 ± 5°C	0–18 & 28–55°C per 1°C
10.00000 V	—	—	±(35 ppm +5 ppm)	±(5 ppm +1 ppm)
1000.000 V	±(20 ppm +6 ppm)	±(35 ppm +10 ppm)	±(45 ppm +10 ppm)	±(5 ppm +1 ppm)

(We do not combine the two sums under square roots above to stress that one sum contains originally symmetric errors and the other – the errors that were originally asymmetric but which have been recomputed to become symmetric.) The absolutely constant elementary error must now be taken into account, and it too can have asymmetric limits. Again, these limits must be summed arithmetically with the limits $\theta_{r,\alpha}$ and $\theta_{l,\alpha}$:

$$\begin{aligned} U_{r,\alpha} &= H_r + \theta_{r,\alpha} \\ U_{l,\alpha} &= H_l + \theta_{l,\alpha} \end{aligned} \tag{4.16}$$

As an example of estimating the inaccuracy of a single measurement under rated conditions, consider the measurement of voltage using, again, a digital multivoltmeter whose errors are rated in Table 4.3. Assume it is known (from other parts of the documentation) that this instrument's indication has six and a half digits: if the seventh, invisible, digit is less than 5, then the sixth digit will not increase whereas if the seventh digit is 5 or greater, the sixth digit will increase by 1. Thus, the random rounding error is limited to half the value of the sixth digit.

Assume the measurement occurs 12 months after the last calibration of the instrument and the voltmeter is used in the range of 10 V. Assume further the voltmeter is mounted in an automated test rack with internal temperature of 32°C and is indicating 5.00135 V. We need to calculate the uncertainty of this measurement.

Using the 12-month specifications, the limits of the intrinsic error of this meter are $(5.00135\,\mathrm{V} \times 35 \times 10^{-6} + 10\,\mathrm{V} \times 5 \times 10^{-6}) = 0.225\,\mathrm{mV}$. Since the instrument works in temperature outside the reference conditions, the temperature coefficient, according to the last column, is $(5.0135\,\mathrm{V} \times 5 \times 10^{-6} + 10\,\mathrm{V} \times 1 \times 10^{-6})$ per 1°C, or $35 \times 10^{-6}\mathrm{V}/°\mathrm{C}$. Thus, with the operating condition being 4°C over 28°C, the additional error is $4 \times 35 \times 10^{-6} = 0.14\,\mathrm{mV}$. The rounding error does not exceed $5 \times 10^{-6}\mathrm{V} = 0.005\,\mathrm{mV}$.

We now combine the elementary errors in two ways. The arithmetic sum of the obtained limits is $\pm(0.225 + 0.14 + 0.005)\,\mathrm{mV} = \pm 0.37\,\mathrm{mV}$. Probabilistic summation according to (4.3) with $\alpha = 0.95$ gives $\pm 1.1 \times 0.265\,\mathrm{mV} = \pm 0.29\,\mathrm{mV}$. Because the probabilistic result is smaller, we should take as uncertainty of the measurement $\pm 0.29\,\mathrm{mV}$ or, after rounding, $\pm 0.3\,\mathrm{mV}$.

Another example of estimating the inaccuracy of a single measurement under rated condition is given in Sect. 8.1.

4.8 Accuracy of Multiple Measurements

Multiple measurements are a classic object of mathematical statistics and the theory of measurement errors. Under certain restrictions on the starting data, mathematical statistics give elegant methods for analyzing observations and for estimating measurement errors. Unfortunately, the restrictions required by mathematics are not often satisfied in practice. Then these methods cannot be used, and practical methods for solving the problems must be developed. But even in this case, the methods of mathematical statistics provide a point of reference and a theoretical foundation.

We previously argued for a position that a multiple direct measurement is in essence a series of repeated single measurements. From this perspective, the inaccuracy of a single measurement comes into fore, and the need to account for it becomes obvious. Thus, our problem is to find an estimate of the measured quantity and the inaccuracy of this estimate. Our starting data comprises inaccuracy of the underlying single measurement, θ_0 or u_0, and the series of the result of repeated single measurements $\{x_i\}$, $i = 1, \ldots, n$.

Usually, the estimate of the measurand is taken as the arithmetic mean of the results of the repeated measurement. As noted previously (Sect. 3.2), this gives an unbiased, consistent, and efficient estimate of the true value of the measured quantity only if the observations, or equivalently the measurement errors, have a normal distribution. In fact, irrespective of the form of the distribution of the measurement errors, the arithmetic mean has three important properties:

1. The sum of the deviations from the arithmetic mean is equal to 0. Let $x_i, \ldots, x_n$ be a group of observations whose arithmetic mean is $\bar{x}$. We construct the differences $x_i - \bar{x}$ for all $i = 1, \ldots, n$ and find their sum:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}.$$

   As both $\sum_{i=1}^{n} x_i = n\bar{x}$ and $\sum_{i=1}^{n} \bar{x} = n\bar{x}$,

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

   This property of the arithmetic mean can be used to check the calculations.
2. The sum of the squares of the deviations from the arithmetic mean is smaller than the sum of the squares of the deviations from any other estimate $\tilde{A}$ of true value A. Consider the function

$$Q = \sum_{i=1}^{n}(x_i - \tilde{A})^2.$$

We shall find $\tilde{A}$ that minimizes Q. To this end, we find

$$\frac{dQ}{d\tilde{A}} = -2\sum_{i=1}^{n}(x_i - \tilde{A})$$

and set it to zero; hence, we obtain

$$\sum_{i=1}^{n}(x^i - \tilde{A}) = 0, \qquad \sum_{i=1}^{n} x_i = n\tilde{A}, \quad \text{and} \quad \tilde{A} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

As $dQ/d\tilde{A} < 0$ if $\tilde{A} < x$ and $dQ/d\tilde{A} < 0$ if $\tilde{A} > x$, the value $\tilde{A} = \bar{x}$ minimizes function Q.

3. According to the central limit theorem, the sum of independent random quantities, regardless of their distribution functions, tends to a normal distribution as the number of the random quantities grows to infinity. Equivalently, the arithmetic mean of independent observations tends to a normal distribution when the number of observations grows to infinity. In practice, a relatively few random quantities lead to a sum that can be viewed as normally distributed. In particular, in the context of measurement accuracy, one can consider the sum – or the arithmetic mean – of five random quantities with uniform distribution function to be normally distributed.

A drawback of the arithmetic mean is its high sensitivity to outlying observations. Another popular estimate of the measurand is the median. The median is less sensitive to the outliers, but it is also less efficient: its variance exceeds the variance of the arithmetic mean. Indeed, let m_* be the sample median and A be the true value of the measured quantity. It is known [19] that m_* has asymptotically normal distribution with mathematical expectation A and standard deviation

$$\sigma\left(m_*\right) = \sqrt{\pi/2} \times \sigma\left(\bar{x}\right) = 1.25\sigma\left(\bar{x}\right),$$

where $\sigma\left(\bar{x}\right)$ is standard deviation of the arithmetic mean. Since the median is a less-efficient estimate, one needs more data to obtain the same confidence interval for the measurement result using the median than arithmetic mean.

Although the arithmetic mean produces the minimum sum of the squares of the deviations, this only means that it is the most efficient estimate of the measured quantity in the class of estimates that are a linear function of the observations. This estimate becomes most efficient among all possible estimates if the errors are distributed normally. For other distributions, as pointed out in Chap. 3, estimates exist that are more efficient.

From now on, we will assume that we use the arithmetic mean for the estimate of the measured quantity:

$$\tilde{A} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{4.17}$$

Because of random errors, the measurement results are also random quantities; if another series of measurements is performed, then the new arithmetic mean obtained will differ somewhat from the previously found estimate. Thus, the arithmetic mean of a set of measurement results is a random quantity. The spread of the arithmetic means is characterized either by the variance of the arithmetic means or by the standard deviation. In accordance with (3.12) and (3.16), they are estimated from the experimental data as follows:

$$S_{\bar{x}}^2 = \frac{1}{n\,(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{or} \quad S_{\bar{x}} = \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n(n-1)}}. \tag{4.18}$$

In addition, it is possible to construct the confidence interval for A for confidence probability α, which is determined by the inequalities

$$\bar{A} - \Psi_\alpha \leq A \leq \tilde{A} + \Psi_\alpha,$$

where $\Psi_\alpha = t_q S_{\bar{x}}$, and t_q is the percentile of Student's distribution for the significance level $q = 1 - \alpha$ and the degree of freedom $\nu = n - 1$ (see Table A.2). Thus, the random error ψ has the limits $\pm\Psi_\alpha$ with the confidence probability α.

We should note that the random error of the single measurement that forms the basis of the multiple measurement is also included into the random error of the multiple measurement. For this reason, the random error of the single measurement is accounted for twice. It would have been better to deduct this error from the error of the multiple measurement, but that would require knowing the random component of the measuring instrument, besides its intrinsic error.

The situation is different in the case of systematic errors. This error of the multiple measurement is the same as the systematic error of its base single measurement.

In the general form, the error of a measurement result has three components:

$$\zeta = \eta + \vartheta + \psi,$$

where η is the absolutely constant error, ϑ is the conditionally constant error, and ψ is the random error. Therefore, the variance of measurement result is

$$V[\zeta] = V[\vartheta] + V[\psi].$$

Note that $V[\zeta]$ has only two terms because $V[\eta] = 0$. The absolutely and conditionally constant errors are determined by the error of the base single measurement, while the random error depends also on the instability of measurement conditions.

Estimates of $V[\vartheta]$ and $V[\psi]$ can be found using formulas (4.5) and (4.18). Denote them S_ϑ^2 and $S_{\bar{x}}^2$. Denote also the estimate of the combined variance S_c^2. Then the combined standard deviation S_c is

$$S_c = \sqrt{S_\vartheta^2 + S_{\bar{x}}^2}. \tag{4.19}$$

Given S_c, the uncertainty of the measurement result could be calculated from the formula

$$u_c = t_c S_c \tag{4.20}$$

if the coefficient t_c was known; unfortunately, this coefficient is unknown. We will now consider how to estimate it.

As the initial data, i.e., the data on the components of the uncertainty are not known accurately, an approximate estimate of the coefficient t_c can be used. In [48], the following formula was proposed for this purpose:

$$t_c = \frac{\Psi_\alpha + \theta_\alpha}{S\bar{x} + S_\vartheta},$$

where θ_α is the confidence interval boundary of the single measurement error and Ψ_α is the confidence limit of the random error ψ of the multiple measurement (determined using Student's distribution as described earlier).

This formula was constructed based on the following considerations. The coefficient t_q, determining the ratio of the confidence limit and the standard deviation of the random error, is determined by Student's distribution and is known. Given estimates for the confidence limit θ_α and standard deviation S_ϑ of the conditionally constant error, we can introduce an analogous coefficient t_ϑ as their ratio:

$$t_\vartheta = \theta_\alpha / S_\vartheta \tag{4.21}$$

It is natural to assume that the coefficient sought t_c is some function of t_q and t_ϑ, and that the computed t_c corresponds to the same confidence probability. If we take a weighted average of t_q and t_ϑ for the weights $S_\vartheta / (S_{\bar{x}} + S_\vartheta)$ and $S_{\bar{x}} / (S_{\bar{x}} + S_\vartheta)$, respectively, for this function, we obtain the proposed formula:

$$t_c \frac{t_q S_{\bar{x}} + t_\vartheta S_\vartheta}{S_{\bar{x}} + S_\vartheta} = \frac{\Psi_\alpha + \theta_\alpha}{S_{\bar{x}} + S_\vartheta}. \tag{4.22}$$

If the base single measurements are performed under rated conditions for the used measuring instrument, then $\theta_\alpha = u_\alpha$. Recall that $u_\alpha = k_\alpha \sqrt{\sum_{i=1}^{n} \theta_i^2}$, and $k_{0.95} \approx 1.1$, and $k_{0.99} \approx 1.4$ for $n < 5$. More accurate values of this coefficient can be found using Table 4.1 and Fig. 4.2. For $n > 5$, one can calculate u_α using (4.6).

Under reference conditions, θ_α is determined by the limits of error θ_0 of the single measurement and the given confidence probability α according to a method illustrated in Fig. 4.3. The figure shows the CDF of error uniformly distributed in $[-\theta_0, +\theta_0]$. The confidence limit for confidence probability α is the quantile θ_α for probability $p = 1-(1-\alpha)/2 = (1+\alpha)/2$. We can compute this quantile by considering two similar triangles highlighted in the figure with dotted lines, one with a side of size $2\theta_0$ and the other with the corresponding side of size $(\theta_0 + \theta_\alpha)$. From

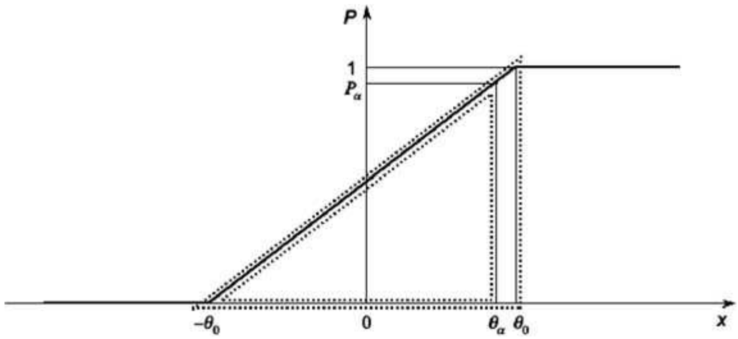


Fig. 4.3 Computing limits of confidence interval for uniformly distributed random error

the similarity of the triangles follows the equality $1/(2\theta_0) = (1+\alpha)/2(\theta_0 + \theta_\alpha)$, which gives

$$\theta_\alpha = \alpha\theta_0. \tag{4.23}$$

To use formula (4.22), its accuracy must be estimated. The extreme cases are those when the conditionally constant systematic error ϑ has a normal or uniform distribution. The distribution of the random error Ψ_α of the arithmetic mean can be assumed to be asymptotically normal.

If both conditionally constant and random errors have a normal distribution, then $t_q = t_\vartheta$, and as follows from formula (4.22), $t_c = t_q$. As the composition of normal distributions gives a normal distribution, the obtained value of t_c is exact.

If the conditionally constant error is uniformly distributed, the results of calculations based on the approximate formula (4.22) must be compared with the results obtained from the exactly constructed composition of normal and uniform distributions. The expression for the distribution density of the composition of centered uniform and normal distributions is known from the theory of probability:

$$f(z) = \frac{1}{2h}\int_{-h}^{h} \frac{1}{\sigma\sqrt{2\pi}} e^{-(z-y)^2/2\sigma^2}\,dy, \tag{4.24}$$

where h is equal to one-half the interval in which the uniform random quantity is distributed and σ is the standard deviation of the normal random quantity.

The variance of this distribution is

$$\sigma_c^2 = \sigma^2 + \frac{h^2}{3} = \sigma^2\left[1 + \frac{1}{3}\left(\frac{h}{\sigma}\right)^2\right]. \tag{4.25}$$

The above distribution depends on both the ratio (h/σ) and σ. We will analyze it for $\sigma = 1$. In addition to simplifying the calculations, this will make the composed distribution universal, in the same way the standard normal distribution is universal. Transforming the density to the probability distribution and setting $\sigma = 1$, we obtain

$$\underset{z>0}{F(z)} = 0.5 + \frac{1}{2h\sqrt{2\pi}} \int_0^z \int_{-h}^{h} e^{-(v-y)^2/2} dy dv. \tag{4.26}$$

The variance of this distribution becomes

$$\sigma_{c,1}^2 = 1 + \frac{1}{3}\left(\frac{h}{\sigma}\right)^2. \tag{4.27}$$

The starting distributions are symmetric relative to 0. Hence, the resulting distribution is also symmetric. For this reason, the limits of the confidence interval corresponding to the probability α are quantile z_p of distribution (4.26) for probability p and quantile z_{1-p} for probability $(1-p)$, where $p = (1-\alpha)/2$. Indeed, $|z_p| = |z_{1-p}|$ because the distribution is symmetrical, and the amount of probability covered by this interval is $1–2p = \alpha$. Because confidence probability α is always taken to be more than 0.5, $p < 0.5$ and therefore quantile z_p gives the left limit and z_{1-p} the right limit of the confidence interval.

Table 4.4 gives values of z_{1-p} calculated using formula (4.26) for confidence probability $\alpha = 0.90$, 0.95, and 0.99. As mentioned above, z_{1-p} represents the exact confidence limit of the combined error that corresponds to $\sigma_{c,1}$. If we instead compute the overall uncertainty $u_{c,1}$ for the same $\sigma_{c,1}$ and confidence probability using formulas (4.22) and (4.20), the relative error introduced by the use of the approximate formula (4.22) will be

$$\delta = \frac{u_{c,1} - z_{1-p}}{z_{1-p}}.$$

Although the above confidence limits were calculated for $\sigma = 1$, it is easy to recompute them for other values of σ. Since the distribution functions for $\sigma \neq 1$ and $\sigma = 1$ differ only in their scaling factor σ_c on the abscise axis, recomputation can be done in a way completely analogous to how one uses quantiles of the standard normal distribution with $\sigma = 1$ to obtain quantiles of normal distributions with $\sigma \neq 1$. Specifically,

$$z_{1-p,\sigma} = \sigma_c z_{1-p}, \tag{4.28}$$

Table 4.4 Quantiles for the composition of centered normal and uniform distributions

h/σ	0.50	1.0	2.0	3.0	4.0	5.0	6.0	8.0	10
$z_{0.95}$ ($\alpha = 0.90$)	1.71	1.90	2.49	3.22	4.00	4.81	5.65	7.34	9.10
$z_{0.975}$ ($\alpha = 0.95$)	2.04	2.25	2.90	3.67	4.49	5.34	6.22	8.00	9.81
$z_{0.995}$ ($\alpha = 0.99$)	2.68	2.94	3.66	4.49	5.36	6.26	7.17	9.02	10.90

where $z_{1-p,\sigma}$ is the quantile of the combined distribution for an arbitrary σ. For example, consider a measurement where $S_{\bar{x}} = 2$ and $\theta_0 = 2$. This corresponds to $\sigma = 2, h = 2$ and $\sigma_c = \sqrt{4 + \frac{4}{3}} = 2.31$. Thus, $h/\sigma = 1$. If we take confidence probability 0.9, we obtain from Table 4.4 the quantile $z_{1-p} = 1.90$ and the quantile $z_{1-p,\sigma} = 4.4$.

Again, the quantile $z_{1-p,\sigma}$ represents the precise value of the confidence limit of the combined error having variance σ_c^2 for confidence probability α. Then, the inaccuracy of the approximate confidence limit u_c in the case of an arbitrary σ becomes:

$$\delta = \frac{u_c - \sigma_c z_{1-p}}{\sigma_c z_{1-p}} = \frac{u_c - z_{1-p,\sigma}}{z_{1-p,\sigma}}. \tag{4.29}$$

To estimate the inaccuracy of formula (4.22) we should contrast the empirical formula (4.20) with the corresponding theoretical formula $z_{1-p,\sigma} = t_r\sigma_c$. The comparison should be done for $S_c = \sigma_c$, bringing (4.20) to the form $u_c = t_c\sigma_c$. Then, by dividing the nominator and denominator of the right-hand side of (4.29) by σ_c, we obtain

$$\delta = \frac{t_c - t_r}{t_r}.$$

Thus, we can analyze the accuracy of (4.22) by considering the accuracy of coefficient t_c relative to its "true value" t_r. We proceed with this analysis next.

We can compute a series of values of coefficient t_r from the data in Table 4.4. These values are presented in Table 4.5, which also gives the corresponding values of $\sigma_{c,1}$.

We shall now compute coefficient t_c using the approximate formula (4.22). The limits of the confidence interval of the conditionally constant error, determined based on the uniform distribution in accordance to (4.23), give θ_α. Because in this case $h = \theta_0$, we have

$$\theta_\alpha = \alpha h.$$

The limit of the confidence interval for the normal distribution with the same confidence probability will be

$$\Psi_\alpha = z_{\frac{1+\alpha}{2}}\sigma,$$

Table 4.5 Values of the combined standard deviation σ_c and of the coefficient t_r as a function of the parameters of the normal and uniform distributions

h/σ	0.5	1	2	3	4	5	6	8	10
$\sigma_{c,1}$	1.04	1.15	1.53	2.00	2.52	3.06	3.51	4.72	5.85
t_r ($\alpha = 0.90$)	1.65	1.64	1.63	1.61	1.59	1.58	1.57	1.56	1.55
t_r ($\alpha = 0.95$)	1.96	1.95	1.90	1.84	1.78	1.75	1.72	1.69	1.67
t_r ($\alpha = 0.99$)	2.57	2.54	2.40	2.24	2.13	2.05	1.99	1.91	1.86

Table 4.6 Values of the coefficient t_c as a function of the parameters of the normal and uniform distributions

h/σ	0.5	1	2	3	4	5	6	8	10
t_{1c} ($\alpha = 0.90$)	1.63	1.61	1.60	1.59	1.58	1.58	1.58	1.57	1.57
t_{2c} ($\alpha = 0.95$)	1.89	1.84	1.79	1.76	1.74	1.73	1.72	1.70	1.69
t_{3c} ($\alpha = 0.99$)	2.38	2.26	2.11	2.03	1.97	1.94	1.91	1.87	1.84

Table 4.7 Deviations of coefficient t_c from t_r (in %)

h/σ	0.5	1	2	3	4	5	6	8	10
δ_1 ($\alpha = 0.90$)	−1.2	−1.9	−1.8	−1.1	−0.6	0.0	0.8	0.6	1.2
δ_2 ($\alpha = 0.95$)	−3.6	−5.5	−5.7	−4.1	−2.2	−1.3	0.0	0.5	1.0
δ_3 ($\alpha = 0.99$)	−7.4	−11.0	−12.1	−9.4	−7.3	−5.5	−4.0	−2.2	−1.1

where $z_{\frac{1+\alpha}{2}}$ is the quantile of the standard normal distribution for probability $\frac{1+\alpha}{2}$. Expression (4.22) assumes the form

$$t_c = \frac{z_{\frac{1+\alpha}{2}}\sigma + \alpha h}{\sigma + h/\sqrt{3}} = \frac{z_{\frac{1+\alpha}{2}} + \alpha\frac{h}{\sigma}}{1 + \frac{h}{\sigma}\sqrt{3}}.$$

The values of t_c, calculated for the same ratios h/σ and confidence probabilities as were used for calculating t_r, are presented in Table 4.6.

We now can compute the errors δ calculated based on the data given in Tables 4.5 and 4.6; these errors are summarized in Table 4.7.

Overall, as Table 4.7 shows, the errors from using the approximate formula are in all cases negative and their absolute magnitude does not exceed 12% for $\alpha = 0.99$, 6% for $\alpha = 0.95$ and 2% for $\alpha = 0.90$. Further, these errors are the highest when h is between σ and 2σ; they decrease for h less than σ or greater than 2σ.

Observe that Table 4.7 lists the inaccuracy of t_c in the extreme case when this inaccuracy is the highest. Moreover, for this case, when one of the component errors is uniformly and the other normally distributed, we have obtained the exact solution, so that the case with the highest inaccuracy can be avoided by using t_r from Table 4.5. But even the worst-case error is acceptable. We would like to repeat that these errors decrease as the distribution of the systematic errors approaches the normal distribution.

In summary, the above scheme presents a general method for estimating the uncertainty of a measurement that contains both random and systematic components. Our analysis (with results summarized in Table 4.7) shows that even in the worst case, when the conditionally constant systematic error is uniformly distributed, this scheme is sufficiently accurate to be used in practice.

4.9 Comparison of Different Methods for Combining Systematic and Random Errors

The above method for combining systematic and random errors is not the only method that has been proposed. In this section, we describe four other methods, compare all the methods on a specific example, and discuss the applicability of these methods and other issues.

1. The US National Institute of Standards and Technology (NIST) in publication [20] presents the following formula (reformulated according to our notation) for combining the component errors (this formula is also mentioned in [6]):

$$u = \theta + \Psi_\alpha, \tag{4.30}$$

where $\theta = \sqrt{\sum_{i=1}^{m} \theta_i^2}$ if $\{\theta_i\}$ $i = 1, \ldots, m$, are independent systematic components, and $\theta = \sum_{i=1}^{m} \theta_i$ if they are dependent, and $\Psi_\alpha = t_q S_{\bar{x}}$.

This method is justified when the absolutely constant elementary errors predominate the overall error. This is often the case in measurements performed in the context of checking and calibrating measuring instruments, which is an area of a particular interest to NIST as an organization. But this method cannot be applied to arbitrary measurements, because in most cases, it results in overestimation of the uncertainty.

It is necessary to note that NIST issued in 1994 Guidelines where the combined uncertainty is calculated in accordance with the method from GUM [2] (which we consider shortly) and not based on formula (4.30).

2. The standard reference [6] and the manual [14] preceding it give two different formulas for calculating the uncertainties with confidence probabilities of 0.95 and 0.99:

$$u_{c,0.99} = \theta + t_{0.95} S_{\bar{x}}, \qquad u_{c,0.95} = \sqrt{\theta^2 + (t_{0.95} S_{\bar{x}})^2}.$$

The coefficient $t_{0.95}$ is chosen according to Student's distribution in both cases for the confidence probability 0.95 ($q = 0.05$) and degrees of freedom $\nu = n - 1$.

The formulas appear to be ad hoc. They are not grounded in probabilistic reasoning, and yet they assign a stated confidence probability of 0.99 or 0.95 to the result.

3. Another method appeared in the Fourth Draft of the Guide to the Expression of Uncertainty in Measurements issued by working group ISO/TAG4/WG3 before the guide itself was published. In this method, the elementary systematic errors are regarded as uniformly distributed random quantities. However, the limit of their sum is calculated with the formula $\theta = \sqrt{\sum_{i=1}^{m} \theta_i^2}$, i.e., without using the indicated model.

The systematic and random errors are combined with a formula that is almost the same as (4.20). The only difference lies in the coefficient t_c. Here, the coefficient is found from Student's distribution corresponding to the selected confidence probability and the effective degrees of freedom ν_{eff}. The following formula is given to calculate ν_{eff}:

$$\frac{S_c^4}{\nu_{\text{eff}}} = \frac{S_{\bar{x}}^4}{\nu} + \sum_{i=1}^{m} \left(\frac{\theta_i^2}{3} \right)^2 .$$

It is assumed here that the random component of uncertainty has a degree of freedom $\nu = n-1$, and each component of the systematic error has a degree of freedom equal to one. However, the notion of a degree of freedom is not applicable to random variables with a fully defined distribution function. Therefore, it is unjustified to assume that a quantity with uniform distribution within given limits has a degree of freedom equal to one (or to any other finite number). Thus, the formula under discussion is not mathematically grounded.

4. GUM [2] presents a method that is similar to the method of the Fourth Draft (and in other drafts), but without the ungrounded computation of coefficient t_c. Instead, GUM assumes t_c to be constant: $t_c' = 2$ for $\alpha = 0.95$ and $t_c'' = 3$ for $\alpha = 0.99$. As we will see later, this method is good if the systematic error is small relative to the random error but can be deficient in other cases.
5. Finally, this book proposes a method with the resulting formulas (4.20) and (4.22).

We shall compare all the methods above using two numerical examples.

Consider a multiple measurement comprising $n = 16$ single measurements. Suppose that as a result of some measurement, the following indicators of its errors were obtained:

$$S_{\bar{x}} = 1, \quad \theta_0 = 3.$$

Suppose also that the random errors have a normal distribution and that the (conditionally constant) systematic errors have a uniform distribution. Then for the exact solution we can take the confidence limits presented in Table 4.4. As usual, we shall take $\alpha_1 = 0.95$ and $\alpha_2 = 0.99$. Then

$$u_{T,0.99} = 4.49, \qquad u_{T,0.95} = 3.67.$$

There is a slight inaccuracy in viewing the above as "exact solution" as we assumed that $S_{\bar{x}} = \sigma_{\bar{x}}$. But for $n = 16$, any discrepancy from this assumption is insignificant, and we shall neglect it.

When applied to this measurement, the considered methods give the following results.

1. The coefficients of Student's distribution with $\nu = n - 1 = 15$ and the indicated values of the confidence probabilities will be as follows:

$$t_{0.99}(15) = 2.95, \qquad t_{0.95}(15) = 2.13,$$
$$\Psi_{0.99} = 2.95 \times 1 = 2.95, \qquad \Psi_{0.95} = 2.13 \times 1 = 2.13.$$

Therefore, $u_{1,0.99} = 3 + 2.95 = 5.95$ and $u_{1,0.95} = 3 + 2.13 = 5.13$.

2. We shall make use of the calculations $t_{0.95}$ and $\Psi_{0.95}$ that were just performed:

$$u_{2,0.99} = 3 + 2.13 \times 1 = 5.13, \qquad u_{2,0.95} = \sqrt{3^2 + (2.13)^2} = 3.68.$$

3. We will need the following values to apply this method,

$$S_{\vartheta}^2 = 9/3 = 3, \qquad S_{\vartheta} = 1.73,$$
$$S_c^2 = 1 + 3 = 4, \qquad S_c = 2.$$

   We shall calculate the effective number of degrees of freedom:

$$\frac{4^2}{\nu_{\text{eff}}} = \frac{1}{15} + 3^2, \qquad \frac{16}{\nu_{\text{eff}}} = 9.07, \qquad \text{and} \qquad \nu_{\text{eff}} = 2.$$

   Next, we find from Student's distribution $t_{3,0.99} = 9.9$ and $t_{3,0.95} = 4.3$. Correspondingly, we obtain

$$u_{3,0.99} = 9.9 \times 2 = 19.8, \qquad u_{3,0.95} = 4.3 \times 2 = 8.6.$$

4. We have, in this case, $S_c = \sqrt{S_{\bar{x}}^2 + S_{\vartheta}^2} = \sqrt{1+3} = 2.0$. Because $t_{0.99} = 3$ and $t_{0.95} = 2$, we obtain

$$u_{4,0.99} = 3.2 = 6, \qquad u_{4,0.95} = 2.2 = 4.$$

5. Formulas (4.20) and (4.22) give $S_{\vartheta} = 1.73$ and $S_c = 2.0$. Then,

$$t_{5,0.99} = \frac{2.95 \times 1 + 0.99 \times 3}{1 + 1.73} = \frac{5.92}{2.73} = 2.17,$$
$$t_{5,0.95} = \frac{2.13 \times 1 + 0.95 \times 3}{1 + 1.73} = \frac{4.98}{2.73} = 1.82,$$
$$u_{5,0.99} = 2.17 \times 2 = 4.34, \qquad u_{5,0.95} = 1.82 \times 2 = 3.64.$$

Let us compare the estimated uncertainties with the exact values $u_{T,0.99}$ and $u_{T,0.95}$. The errors of these computations as compared to the exact values are summarized in Table 4.8. Furthermore, Table 4.9 presents these errors for the case $\theta_0 = 0.5$ and the same values $S_{\bar{x}} = 1$ and $n = 16$, calculated similarly.

In comparison with the previous example, method 4 and especially method 3 in this case show a significant decrease in error. It is not surprising because now the systematic component is insignificant relative to the random component.

We can make the following observations from these examples:

1. As expected, method 3 cannot be used when the systematic error is significant, as in the first example. This method shows a significant decrease in error in the second example, where the systematic component is relatively small.

Table 4.8 Errors of different methods of uncertainty calculation for the case where $\theta_0 = 3,\ S_{\bar{x}} = 1,\ n = 16$

Method of computation	$(u_i - u_T)/u_T \times 100\%$	
	$\alpha = 0.99$	$\alpha = 0.95$
1	32	39.0
2	14	0.3
3	340	132.0
4	34	6.0
5	3	0.8

Table 4.9 Errors of different methods of uncertainty calculation, for the case where $\theta_0 = 0.5,\ S_{\bar{x}} = 1,\ n = 16$

Method of computation	$(u_i - u_T)/u_T \times 100\%$	
	$\alpha = 0.99$	$\alpha = 0.95$
1	29	30
2	2	7
3	13	8
4	12	2
5	4	3

2. Method 2, irrespective of the remarks made earlier, gave satisfactory results in both examples.
3. Method 1, as expected, produced estimates that were too high in both examples.
4. Method 4 is good if the systematic component is small relative to the random component.
5. Our proposed method 5 gave the best results in both examples.

Examples are not, of course, proof, but they nonetheless illustrate well the considerations stated earlier.

Chapter 5
Indirect Measurements

5.1 Terminology and Classification

As introduced in Chap. 1, *indirect measurement* is a measurement in which the value of the unknown quantity sought is calculated using measurements of other quantities related to the measurand by some known relation. These other quantities are called *measurement arguments* or, briefly, *arguments*.

In an indirect measurement, the true value of a measurand A is related to the true values of arguments A_j $(j = 1, \ldots, N)$ by a known function f. This relationship can be represented in a general form as

$$A = f(A_1, \cdots, A_N). \tag{5.1}$$

This equation is called a *measurement equation*. The specific forms of measurement equations can be considered as mathematical models of specific indirect measurements.

Various classifications of indirect measurement are possible. We shall limit ourselves to classifications that will be useful for our purposes.

From the perspective of conducting a measurement, we shall distinguish *single* and *multiple indirect measurements*. In single measurements, all arguments are measured once. In a multiple measurement, all arguments are measured several times.

Multiple indirect measurements differ in a subtle but important way from multiple direct measurements. Whereas the latter involves obtaining a measurand estimate in every constituent single measurement and then processing these estimates to obtain the overall measurement result, the former typically involves estimating arguments from the corresponding multiple argument measurements and then obtaining the overall indirect measurement result (except for the method of reduction considered later in this chapter). Thus, the indirect measurement itself is not repeated: the estimate of the measurand is obtained once all argument measurements are completed. This is why, unlike direct measurements, single indirect measurements cannot be considered as a base form of multiple indirect measurements.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_5, © Springer Science+Business Media, LLC 2010

According to the form of the functional dependency (5.1), we shall distinguish *linear* and *nonlinear indirect measurements.* In the case of a linear indirect measurement, the measurement equation has the form

$$A = b_0 + \sum_{j=1}^{N} b_j A_j, \tag{5.2}$$

where $\{b_j\}(j = 0, \ldots, N)$ are constant coefficients. Nonlinear indirect measurements are diverse, and therefore, it is impossible to represent all of them with one general form of measurement equation.

The physics of the processes of indirect measurements gives us another important classification criterion. To motivate this classification, let us compare the accurate measurement of the density of a solid with the measurement of the temperature coefficient of the electrical resistance of a resistor.

To measure the density of a solid, its mass and volume should be measured independently, with consistent accuracy. In the temperature coefficient measurement, the resistance of the resistor and its temperature are measured simultaneously, which means that the measurements of these arguments are not independent. Thus, we shall distinguish *dependent* and *independent indirect measurements*.

Indirect measurements, like any measurements, are divided into static and dynamic. Recall that we call a measurement dynamic if it utilizes a measuring instrument in the dynamic regime [51]. According to this definition, a multiple indirect measurement should be considered dynamic if any of its arguments are measured with instruments in the dynamic regime. Such measurements are theoretically possible but hardly encountered in practice. For this reason, multiple indirect measurements are usually static; only single indirect measurements can be either static or dynamic.

5.2 Correlation Coefficient and Its Calculation

The traditional methods for estimating the uncertainty of indirect measurements include the calculation of the correlation coefficient.

Later in this book, we shall develop a new theory, which obviates any need for the correlation coefficient. However, given the traditional importance of the correlation coefficient and a great deal of confusion in metrology with its calculation, it makes sense to describe here a clear procedure for calculation of the correlation coefficient.[1]

The mathematical foundation and methods of the correlation coefficient calculations can be found in many books on the probability theory and mathematical

[1] I agree with R.H. Dieck that "probably one of the most misunderstood and misused statistics is the correlation coefficient" [22].

statistics, for example, [53]. Consider two random quantities X and Y with mathematical expectations equal to zero and finite variances. Denote their variances as $V[X] = \sigma_X^2$ and $V[Y] = \sigma_Y^2$.

The variance of a random quantity $Z = X + Y$ can be calculated using the equation

$$V[Z] = E[((X+Y) - E[X+Y])^2] = E[(X+Y)^2] = E[X^2]+E[Y^2]+2E[XY]. \tag{5.3}$$

The last term $E[XY]$ is named *second mixed moment or covariance.*

The covariance divided by the square root of the product of variances $\sigma_X^2\sigma_Y^2$ gives the correlation coefficient ρ_{XY}:

$$\rho_{XY} = \frac{E[XY]}{\sigma_X \sigma_Y}.$$

The value of the correlation coefficient always lies within [–1, +1], and if $|\rho_{XY}| = 1$, then there is a linear functional dependency between X and Y. When $\rho_{XY} = 0$, X and Y are uncorrelated, although it does not mean that they are independent. Otherwise, when $0 < |\rho_{XY}| < 1$, the nature of the dependency between X and Y cannot be determined unambiguously: It can be stochastic as well as functional nonlinear dependency. Therefore, in the last case, if the knowledge about the nature of the dependency between X and Y is required, it can only be obtained based on physical properties of the problem rather than inferred mathematically.

From the above formulas, we obtain

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y. \tag{5.4}$$

In practice, we have to work not with the exact values of parameters of random quantities but with their estimates. So, instead of variances $\sigma_Z^2, \sigma_X^2, \sigma_Y^2$, and the correlation coefficient ρ_{XY}, we have to use their estimates S_Z^2, S_X^2, S_Y^2 (we will also use interchangeably $S^2\,(X)$ to denote an estimate of the variance of random quantity X) and r_{XY}. If n is the number of measured pairs $(x_i,\ y_i)$ of random quantities X and Y $(i = 1, \ldots, n)$, and $\bar{x}$ and $\bar{y}$ are averages over n observed values of X and Y, then

$$S_x^2 = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n-1}, \quad S_y^2 = \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2}{n-1}.$$

The estimate of $E[XY]$, which we denote as m_{XY}, will be

$$m_{XY} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

Then, $r_{XY} = m_{XY}/S_X\,S_Y$.

Thus, the calculation formulas for the correlation coefficient of two random quantities and the variance of their sum are as follows:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_X S_Y}, \tag{5.5}$$

$$S_Z^2 = S_X^2 + S_Y^2 + 2r_{XY}S_X S_Y. \tag{5.6}$$

The estimates of the variances of the average values $\bar{x}$ and $\bar{y}$ are known to be

$$S_{\bar{x}}^2 = \frac{S_X^2}{n} \quad \text{and} \quad S_{\bar{y}}^2 = \frac{S_Y^2}{n}.$$

Then, by dividing (5.6) by n, we obtain the estimate of the variance of the mean value of Z:

$$S_{\bar{Z}}^2 = S_{\bar{x}}^2 + S_{\bar{y}}^2 + 2r_{XY}S_{\bar{x}}S_{\bar{y}}. \tag{5.7}$$

The correlation coefficient estimation here is the same as in (5.5). One can also use $S_{\bar{x}}$ and $S_{\bar{y}}$ for the calculation of the correlation coefficient estimation using the fact that $S_X S_Y = nS_{\bar{x}}S_{\bar{y}}$. Then, (5.5) will change to the following:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n(n-1)S_{\bar{x}}S_{\bar{y}}}. \tag{5.8}$$

It is necessary to stress that, in order to compute the correlation coefficient between random quantities X and Y, the number of realizations of X and Y (e.g., the number of measurements of X and Y) must be the same. Moreover, each pair of these realizations must be obtained under the same conditions, for example, at the same time, at the same temperature, and so on.

The theory of correlations says that realizations x_i and y_i must belong to the same event i. A clear illustration of this statement is given by the classic example of the accuracy analysis of firing practice. Here, each event is one shot. Each shot i is described by a pair of values x_i and y_i that express the deviation of the bullet from the center of the target in orthogonal coordinates. In the case of an indirect measurement, one event is the set of matched measurement results of all arguments. This event corresponds to a point in the multidimensional space with arguments as coordinates. We shall call this point a *measurement vector*.

In the above-mentioned example of the measurement of the temperature coefficient of the electrical resistance of a resistor, each pair of measurements of the resistance and temperature is a measurement vector.

5.3 The Traditional Method of Experimental Data Processing

The processing of experimental data obtained in a measurement consists of two steps. In the first step, we estimate the value of the measurand, and in the second step, we calculate the inaccuracy of this estimate.

In an indirect measurement, the first step traditionally is based on the assumption that the estimate $\tilde{A}$ of the measurand A can be obtained by substitution of $\tilde{A}_j$ for A_j in (5.1):

$$\tilde{A} = f\left(\tilde{A}_{1,\ldots,}\tilde{A}_N\right). \tag{5.9}$$

The second step is commonly solved by expansion of the function (5.1) in a Taylor series. Usually the Taylor series is written in the form of an approximate value of the given function, which is brought to its true value with the help of corrections. We want, however, to work with errors rather than with corrections. Thus, we shall therefore write the series in such a form that the approximate value of the function is expressed by adding something to its true value. To simplify further calculation, suppose that the number of arguments $N = 2$. Then, we have the Taylor series in the form:

$$\begin{aligned} f(\tilde{A}_1, \tilde{A}_2) = f(A_1, A_2) &+ \left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right) f(A_1, A_2) \\ &+ \frac{1}{2!}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^2 f(A_1, A_2) + \cdots \\ &+ \frac{1}{m!}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^m f(A_1, A_2) + R_{m+1}, \end{aligned} \tag{5.10}$$

where $\tilde{A}_1 = A_1 + \varsigma_1$, $\tilde{A}_2 = A_2 + \varsigma_2$ (ζ_1 and ζ_2, the errors of $\tilde{A}_1$ and $\tilde{A}_2$), R_{m+1} is the remainder term, and partial derivatives are computed at the point $\left(\tilde{A}_1, \tilde{A}_2\right)$.

The remainder term can be expressed in the Lagrange form:

$$R_{m+1} = \frac{1}{(m+1)!}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^{m+1} f\left(A_1 + \nu_1\zeta_1, A_2 + \nu_2\zeta_2\right), \tag{5.11}$$

where $0 < \nu_{1,2} < 1$.

If the indirect measurement is linear, all terms, except the linear one, are equal to zero.

The general form of the error of an indirect measurement is

$$\zeta = \tilde{A} - A = f(\tilde{A}_1, \tilde{A}_2) - f(A_1, A_2).$$

Turning to the Taylor series, one obtains

$$\zeta = \left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right) f(A_1, A_2)$$
$$+ \frac{1}{2}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^2 f(A_1, A_2) + \cdots + R_{m+1}. \qquad (5.12)$$

In practice, however, only the first linear term is used for error calculations:

$$\zeta = \frac{\partial f}{\partial A_1}\zeta_1 + \frac{\partial f}{\partial A_2}\zeta_2.$$

We will call the partial derivatives above *argument influence coefficients* (not to be confused with influence quantities and coefficients considered in measurements under rated conditions). We shall denote them as follows:

$$w_j = \frac{\partial f}{\partial A_j}, \qquad j = 1, \ldots, N.$$

Now the above equation can be written in the general form:

$$\zeta = \sum_{j=1}^{N} w_j \zeta_j. \qquad (5.13)$$

We emphasize again that all partial derivatives are calculated at the estimates point $(\tilde{A}_1, \tilde{A}_2)$ because the true values A_1, A_2 are unknown.

Putting aside for now absolutely constant errors, we can write

$$\zeta_j = \vartheta_j + \psi_j,$$

where ϑ_j and ψ_j are conditionally constant and random components of the error, respectively. So, (5.13) takes the form:

$$\zeta = \sum_{j=1}^{N} w_j \vartheta_j + \sum_{j=1}^{N} w_j \psi_j. \qquad (5.14)$$

The last formula says that, in indirect measurements, not only the systematic error consists of components, but so also does the random error.

The traditional method considers the random errors only, which means there are no systematic errors in the argument estimation, that is, that $E[\zeta_1] = 0$ and $E[\zeta_2] = 0$. (A method capable of accounting for systematic errors is considered later, in Sect. 5.7.)

The most important characteristic of a random error is its variance. In accordance with the mathematical definition of the variance, we obtain from (5.13), for $N = 2$,

$$v[\zeta] = E\left[(w_1\zeta_1 + w_2\zeta_2)^2\right] = w_1^2 E\left[\zeta_1^2\right] + w_2^2 E\left[\zeta_2^2\right] + 2w_1 w_2 E\left[\zeta_1 \times \zeta_2\right].$$

This formula is different from (5.3) only in the notations. Therefore, one can write

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho_{1,2} w_1 w_2 \sigma_1\sigma_2, \tag{5.15}$$

where

$$\begin{aligned} \sigma^2 &= V[\zeta] = E\left[\zeta^2\right], \quad \sigma_1^2 = E\left[\zeta_1^2\right], \\ \sigma_2^2 &= E\left[\zeta_2^2\right], \quad \text{and} \quad \rho_{1,2} = \frac{E\left[\zeta_1 \times \zeta_2\right]}{\sigma_1\sigma_2}. \end{aligned}$$

We should like to point out that the variance of a random error of the measurement result is identical to the variance of the measurement result:

$$V[\zeta] = V[\tilde{A}].$$

Also note that (5.15) has three terms, which correspond to the case when $N = 2$. When $N = 3$, we shall have six terms. So, with the number of arguments increasing, the complexity of calculations increases rapidly.

In (5.15), the values of variance σ_j^2 and correlation coefficient $\rho_{k,l}$ are unknown and, in practice, their estimations S_j^2 and $r_{k,l}$ are used instead. Taking into account this substitution and assuming the general case of N arguments, (5.15) becomes

$$S^2 = \sum_{j=1}^{N} w_j^2 S^2\left(\tilde{A}_j\right) + 2\sum_{k<l} r_{k,l} w_k w_l S\left(\tilde{A}_k\right) S\left(\tilde{A}_l\right). \tag{5.16}$$

For estimating the variance of the estimation of an argument and correlation coefficient between pairs of arguments, we have the formulas

$$\left.\begin{aligned} S_j^2 = S^2(\tilde{A}_j) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2, \\ r_{k,l} &= \frac{\sum_{i=1}^{n} (x_{ki} - \bar{x}_k)(x_{li} - \bar{x}_l)}{n(n-1)S(\tilde{A}_k)S(\tilde{A}_l)}. \end{aligned}\right\}$$

Here, n is the number of measurement vectors, and x_{ki} is the realization of argument A_k from measurement vector i. In particular, in the formula for the correlation coefficient, the fact that realizations x_{ki} and x_{li} have the same subscript i means that these realizations must be taken from the same vector i. Having the estimates S_j^2 and $r_{k,l}$, one can use (5.16) to obtain the estimate of variance S^2.

If measurements of all arguments are independent, i.e., $\rho_{k,l} = 0$, then (5.16) is simplified:

$$S^2 = \sum_{j=1}^{N} w_j^2 S^2\left(\tilde{A}_j\right). \tag{5.17}$$

This equation gives the following expression for the standard deviation:

$$S = \sqrt{w_1^2 S^2\left(\tilde{A}_1\right) + \cdots + w_N^2 S^2\left(\tilde{A}_N\right)}. \tag{5.18}$$

The last two formulas are often called the *error propagation formulas*, although in reality they express the propagation of variances.

Although (5.18) was derived for the random errors only, it has a wide use as universal formula for the summation of all kinds of errors. This way of error calculation even has a specific name: the *square-root sum method*.

The next problem is to calculate the confidence interval for the true value of the measurand, and hence the uncertainty of the measurement. Within the framework of traditional methods, this problem can only be solved in a mathematically grounded way for linear indirect measurements, although even in this case, the solution is only approximate. For nonlinear measurements, this problem can be solved by linearization of the measurement equation, which leads to additional inaccuracy. However, for dependent indirect measurements the traditional method does not provide any solution, because in this case it is impossible to obtain the probability distribution of the measurement error and to find the appropriate number of degrees of freedom.

Let us consider this simplest case of a linear indirect measurement with normally distributed argument errors. In this case, in principle, one could use Student's distribution, but an exact expression for the degrees of freedom is not known. An approximate solution, which gives an estimate of the degrees of freedom, called the effective number of degrees of freedom, is given by the well-known Welch–Satterthwaite formula [6]:

$$\nu_{\text{eff}} = \frac{\left(\sum\limits_{j=1}^{N} b_j^2 S^2\left(\tilde{A}_j\right)\right)^2}{\sum\limits_{j=1}^{N} \frac{b_j^4 S^4(\tilde{A}_j)}{\nu_j}}, \tag{5.19}$$

where ν_j is the number of degrees of freedom for argument A_j, determined by the number of measurements n_j of A_j: $\nu_j = n_j - 1$. The uncertainty in this case can be calculated as

$$u_c = t_q S,$$

where t_q is found from Student's distribution table for the degrees of freedom ν_{eff} and the significance level $q = 1 - \alpha$ (recall that α is the chosen confidence probability). The obtained uncertainty is approximate because, not knowing the actual degree of freedom, we used its estimate – the effective degrees of freedom.

For nonlinear independent indirect measurements, as already mentioned, the problem of constructing confidence intervals can be solved using linearization of the measurement equation. Linearization is done using the expansion of the measurement equation into Taylor series. In this method, one estimates the standard deviation of the measurement result using (5.18), computes the effective degree of freedom from (5.19) (replacing coefficients b_j with w_j), and then finds the quantile of Student's distribution corresponding to the just-found degree of freedom and chosen confidence probability. Having obtained the quantile, one can calculate the confidence interval for the measurement result, that is, the measurement uncertainty in the same way as for a linear indirect measurement.

This solution, as is the case with linear indirect measurements, is approximate. But its more significant drawback is that it retains only the first, linear, term in the Taylor series. Therefore, the probability distribution of the result of the indirect measurement is unknown and thus the confidence probability assigned to the measurement uncertainty is unlikely valid.

In practice, instead of linearization, the uncertainty is often calculated simply by summation of measurement uncertainties of the arguments using the following formula, which is based on (5.18):

$$u_t = \sqrt{\sum_{j=1}^{N} w_j^2 u_j^2}, \tag{5.20}$$

where u_j is the uncertainty of the measurement of jth argument and w_j is its influence coefficient. Along with (5.18), formula (5.20) is also often called the square-root sum formula. But this square-root sum formula is correct for summing variances, not confidence intervals or uncertainties, and it is unclear if one can call the result a confidence interval or uncertainty.

The next problem is how to calculate the systematic error of an indirect measurement result, and how to combine it with the random error to obtain the overall uncertainty of the indirect measurement result. A reasonable solution of this problem will be discussed below in Sect. 5.7.

5.4 Merits and Shortcomings of the Traditional Method

The traditional method has been used in measurement practice for a long time. It is based on the Taylor series expansion, which allows one to transform input data of an indirect measurement (data obtained from arguments' measurements) into output data, that is, the data about the measurand. This method is universal but, as the analysis presented in [44, 45] showed, it has a number of shortcomings.

First, for a nonlinear function f

$$E[f(X_1, \ldots, X_N)] \neq f(E[X_1], \ldots, E[X_N]),$$

where $X_1, \ldots, X_N$ are random quantities. Therefore, the estimate of the measurand given by (5.9) is incorrect when the measurement equation is nonlinear. Let us evaluate this incorrectness.

Go back to (5.10) and now retain not only the first term but the second one also. Again, assuming $N = 2$ for simplicity, we get

$$\zeta = \left(\frac{\partial f}{\partial A_1}\zeta_1 + \frac{\partial f}{\partial A_2}\zeta_2\right) + \frac{1}{2}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^2 f(A_1, A_2).$$

Assume, as before, ζ_1 and ζ_2 to be free from systematic errors: $E[\zeta_1] = 0$ and $E[\zeta_2] = 0$. Then, the mathematical expectation of the first term is equal to zero:

$$E\left[\left(\frac{\partial f}{\partial A_1}\zeta_1 + \frac{\partial f}{\partial A_2}\zeta_2\right)\right] = w_1 E[\zeta_1] + w_2 E[\zeta_2] = 0.$$

But the variances of the errors ζ_1 and ζ_2 are

$$V[\zeta_1] = \sigma_1^2 > 0 \quad \text{and} \quad V[\zeta_2] = \sigma_2^2 > 0,$$

and therefore the mathematical expectation of the second term in the above Taylor series is not equal to zero. Indeed,

$$\begin{aligned} E[\zeta] &= E\left[\frac{1}{2}\left(\frac{\partial}{\partial A_1}\zeta_1 + \frac{\partial}{\partial A_2}\zeta_2\right)^2 f(A_1, A_2)\right] \\ &= \frac{1}{2}\frac{\partial^2 f}{\partial A_1^2}E[\zeta_1^2] + \frac{1}{2}\frac{\partial^2 f}{\partial A_2^2}E[\zeta_2^2] + \frac{\partial f}{\partial A_1}\cdot\frac{\partial f}{\partial A_2}E[\zeta_1 \times \zeta_2] \\ &= \frac{1}{2}\frac{\partial^2 f}{\partial A_1^2}\sigma_1^2 + \frac{1}{2}\frac{\partial^2 f}{\partial A_2^2}\sigma_2^2 + \frac{\partial f}{\partial A_1}\cdot\frac{\partial f}{\partial A_2}\rho_{1,2}\sigma_1\sigma_2. \end{aligned} \tag{5.21}$$

As $\sigma_1^2 > 0, \sigma_2^2 > 0$ and $|\rho_{1,2}| < 1$, $E[\zeta] = B \neq 0$.

Thus, for nonlinear indirect measurements, the estimate of the measurand given by the traditional method is biased! The bias of the measurement result can be reduced by correction C:

$$C = -B.$$

But even after correction, the estimate of a measurand will not be exact because it takes into account only two terms, whereas the Taylor series may have an infinite number of terms.

This is the first deficiency of the traditional theory of indirect measurements.

It must be considered as an essential disadvantage for it affects the results of measurements.

The second deficiency is that the estimate of the variance of the measurement result, given by (5.16), is imperfect because it is derived using only one linear term in the Taylor series. In other words, the traditional method does not use all of the information contained in the results of measurements of arguments.

The next disadvantage of the traditional method is the problem of the confidence intervals. As we already mentioned, this method does not provide a grounded foundation for constructing the confidence intervals in the case of dependent indirect measurements because in this case it is impossible to obtain the probability distribution of the measurement error and to find the appropriate number of degrees of freedom.

A further drawback is the above-mentioned problem of estimating correlation coefficients that are an inherent part of the traditional method.

As we mentioned earlier, the traditional method allows one to construct a confidence interval for independent indirect measurements. In fact, for nonlinear independent indirect measurements, the most commonly used method is not the method of linearization but the method using the square-root sum formula. However, the justification of applying the square-root formula in this situation has not been proven. Let us investigate this question.

Consider two samples of independent observations of a measurand, each of size n, from the same normal distribution. Let the estimates of their variances be S_1^2 and S_2^2. The confidence limits of the true value of the measurand, according to Student's distribution are

$$u_1 = t_{n-1}S_1 \text{ and } u_2 = t_{n-1}S_2.$$

Coefficient t_{n-1} for both samples is the same since they have the same degree of freedom and the same confidence probability. Let us now combine these samples into one. The combined sample is also from the same normal distribution but with $2n$ observations. The estimate of variance of this sample is

$$S_0^2 = S_1^2 + S_2^2,$$

and the confidence limit is

$$u_0 = t_{2n-1}S_0.$$

Compare the above confidence limit with the one obtained from (5.20):

$$u_0' = \sqrt{u_1^2 + u_2^2} = t_{n-1}S_0.$$

Obviously, $u_0 \neq u_0'$. Let us further look at how big the difference between the two can be. For $n = 10$ and confidence probability $\alpha = 0.95$, we have $u_0' = 2.26 \times S_0$ and $u_0 = 2.10 \times S_0$. Thus, in this case, (5.20) exaggerates the inaccuracy by 8%. We can find in a similar way that with $n = 10$ and three arguments, the difference will be 11%, and with four arguments, 13%. For $n = 5$ and two arguments, the difference reaches 25% and for four arguments, 35%. When $n = 20$, the inaccuracy of (5.20) is 5% and does not depend on the number of arguments.

We can conclude that using (5.20) can be generally acceptable when the number of measurements of each argument is around 10 or more. At the same time, the above analysis reveals several rules one should follow in using (5.20). First, one must keep in mind that this formula exaggerates the uncertainty of the measurement, and the fewer the number of argument measurements the greater the amount of overestimation. Second, to use this formula, one must make sure that measurement uncertainty of each argument has the same degree of freedom. In other words, each argument must be measured the same number of times. Finally, the measurement uncertainty of every argument must be computed for the same confidence probability.

The above analysis also suggests a possibility of introducing a corrective factor $W_t = t_{2n-1}/t_{n-1}$. In the particular example we considered,

$$W_t = t_{2n-1}/t_{n-1} = 2.10/2.26 = 0.93.$$

However, an important point to keep in mind is that the entire analysis is conducted for the case when argument measurement errors are normally distributed. Generalizing the above analysis, a natural suggestion would be to use (5.18) in place of (5.20) for the estimate of combined standard deviation, and then use Student's distribution to build the confidence interval. The degree of freedom in this case is, as we have seen, $\nu = 2\,(n-1)$.

In summary, both methods – linearization and square-root sum – of calculating the uncertainty (i.e., confidence intervals) of independent indirect measurements are approximate. The premise behind these calculations, which is that errors of measurements of the arguments are normally distributed, often does not hold. And only because confidence intervals based on Student's distribution are not highly sensitive to the shape of the distribution functions, these intervals are satisfactory in practice.

Nonetheless, as the analysis of Sects. 5.3 and 5.4 showed, the traditional method and the square-root sum formula (with the enhancements discussed above) still allow one to estimate the uncertainty of independent indirect measurements assuming that the conditions for the applicability of this formula we established do hold.

5.5 The Method of Reduction

As we discussed above, the traditional method of experimental data processing allows one to estimate the uncertainty of the measurement result for independent indirect measurements. But for dependent indirect measurements, this problem remained unsolved. For this reason, in measurements in physics, chemistry, and other scientific disciplines, the uncertainty of a measurement result is taken to be not a confidence interval but the standard deviation. The following *method of reduction* fully solves this problem [34, 44, 46].

Assume that $x_{1i}, x_{2i}, \ldots, x_{Ni}$ are measurement results of arguments from a measurement vector i. Recall that a measurement vector compiles measurements of all arguments performed under the same conditions and at the same time. Each dependent indirect measurement always consists of a definite number of measurement vectors.

So, let n be the number of measurement vectors obtained. These vectors can be represented as a set:

$$\{x_{1i}, x_{2i}, \ldots, x_{Ni}\}, i = 1, \ldots, n.$$

Substituting the results from the ith vector into the measurement equation, we obtain the ith value of the measurand. Denote it by y_i. This transformation is obviously justified because it reflects the physical relationship between the measurand and the measurement arguments.

In the same way, n measurement vectors give us a set of n values of the measurand:

$$\{y_i\}, i = 1, \ldots, n.$$

This set does not differ from a set of data obtained by direct measurements of the measurand A. Hence, we can now use all simple and well-understood methods of direct measurements, which immediately provides an estimate of the measurand

$$\tilde{A} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{5.22}$$

and an estimate of the variance

$$S^2\left(\tilde{A}\right) = \frac{1}{n\,(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2. \tag{5.23}$$

The method of reduction also solves the problem of the calculation of confidence intervals, because we now have the set of n values of the measurand. The confidence limits and therefore the uncertainty of the measurement result are

$$u = t_q S\left(\tilde{A}\right), \tag{5.24}$$

where t_q is found from Student's distribution for the chosen confidence probability and the exact number of degrees of freedom obtained, $\nu = n - 1$.

One might think that the method of reduction imposes special requirements for performing the measurement, namely that the measurements of arguments be performed so that the results can be represented as a number of measurement vectors. However, the traditional method imposes this requirement as well. Indeed, if we have a dependent indirect measurement, all arguments must be measured under the same conditions for the traditional method also, because, otherwise, it is impossible to calculate the correlation coefficients and therefore impossible to estimate the variance of the measurement result.

Thus, the method of reduction has some important advantages over the traditional method:

1. It produces an unbiased estimate of the measurand.
2. It uses all of the information obtained in the course of the measurement.

3. It gets rid of the correlation coefficient in the measurement uncertainty calculations.
4. It gives the exact degree of freedom and allows one to calculate the confidence intervals for the true value of the measurand.

These advantages lead us to conclude that the method of reduction is the preferable method for all kinds of dependent indirect measurements.

It is important to emphasize here that data processing in independent indirect measurements does not require correlation coefficients. As the method of reduction eliminates the need for correlation coefficients in the case of dependent indirect measurements, *the concept of the correlation coefficient is no longer necessary in measurement data processing*.

To conclude, I would like to note that I first proposed this method of reduction approximately in 1970. It found immediate application in national and international comparisons of standards of unit radium mass and in measurements of other radioactive quantities carried at All-Union State Research Institute of Metrology named under Mendeleev in the former Soviet Union. With the reports of these measurements, the information about the method of reduction spread outside that Institute and outside the country. The first formal publication describing this method appeared in 1975 [34]. By now this method has became well known; it is mentioned in the GUM [2] under the name "Approach 2." However, while containing a note that this approach is preferable to "Approach 1" (which is the traditional method), GUM does not explain what the advantages of Approach 2 are.

5.6 The Method of Transformation

The method of reduction described in Sect. 5.5 replaces the traditional method for processing data obtained from dependent indirect measurements. Unfortunately, this method is inapplicable to independent indirect measurements, because it is unclear how to group argument measurements into measurement vectors. The traditional method is applicable but has several drawbacks. In the case of a nonlinear measurement equation, the traditional method involves linearization of the equation, which entails some loss of information obtained from the measurement. The traditional method combines measurement errors of the arguments under the assumption that these errors are all normally distributed. The traditional method used to suffer from the general problem of indirect measurements, namely, the lack of a grounded method for combining the random and systematic errors. While the methods presented in this book allow one to combine these errors, the uncertainty of the result is calculated using an approximate estimate of the degree of freedom leading to loss of accuracy in these calculations.

Consequently, we are presenting a new method for independent indirect measurements, which we call the *method of transformation*. As we will see, this method reduces and in some cases eliminates most of the above drawbacks. The essence of the method of transformation can be understood intuitively if one considers a black box with the arguments as its input and values of the measurand as its output. If we applied fixed argument estimates of all but one argument to the black box's inputs, the black box would transform every observed value of the remaining argument into the corresponding value of the measurand, producing a group of measurement data. Using each argument in this manner, we can obtain a set of these groups, which together provide the basis for the estimate of the measurand along with its uncertainty for a chosen confidence probability.

Turning to a more detailed description, let $A_j,\ j = 1,\ldots,N$ be the arguments of an independent indirect measurement of a measurand x:

$$x = f(A_1,\ldots,A_N). \tag{5.25}$$

We will consider the case when function f in above equation can be separated into multiplicative terms, each depending on one argument[2]:

$$x = f\left(A_1,\ldots,A_2\right) = f_1\left(A_1\right)\cdot\ldots\cdot f_N\left(A_N\right). \tag{5.26}$$

Assume that all arguments but one in (5.26) are fixed to certain values. Let A_d be the remaining argument and let $A_{d,i}\,(i = 1,\ldots,n_d)$ be its observed values. Each value $A_{d,i}$ of the variable A_d, together with the fixed values of all other arguments, produces one value of the measurand. Thus, (5.25) can be presented in the new form

$$x_{d,i} = f_1\left(A_1\right)\cdot\ldots\cdot f_{d-1}\left(A_{d-1}\right)\cdot f_d\left(A_{d,i}\right)\cdot f_{d+1}\left(A_{d+1}\right)\cdot\ldots\cdot f_N\left(A_N\right), \quad i = 1,\ldots,n_d.$$

This formula can be written also as a product of two functions,

$$x_{d,i} = \Psi_d\left(A_1,\ldots,A_{d-1},A_{d+1},\ldots,A_N\right) f_d\left(A_{d,i}\right),$$

where function f_d depends only on the measurement data of argument A_d and function Ψ_d depends only on the chosen values of the remaining arguments.

It will be convenient to rewrite the above as

$$x_{d,i} = C_d f_d\left(A_{d,i}\right), \tag{5.27}$$

where

$$C_d = \Psi_d(A_j), \quad j \neq d. \tag{5.28}$$

The term C_d is determined by values of $A_j\,(j \neq \mathrm{d})$ and therefore is the same for all values $A_{d,i}$. It is called the *transformation coefficient* for argument A_d. We use the estimate of arguments $A_j\,(j \neq d)$, that is, the means $\bar{A}_j$ of their measurements, in

---

[2] One should be able to apply the ideas described here to the case with additive terms as well, although the specific formulas will change.

(5.28) to compute an *estimate* of C_d, $\tilde{C}_d$. Then, using (5.27), a set of n_d measurements of A_d is transformed into the set of the corresponding values of the measurand $\{x_{d,i}\}$, $i = 1, \ldots, n_d$. The same calculations are performed for each argument, producing N sets of values of the measurand. We call the argument A_d used to produce the corresponding group of the measurand data $\{x_{d,i}\}$ the *deriving argument*, and the rest of the arguments *nonderiving arguments.* The groups of argument measurement data are called the *input groups*; the derived groups of the measurement data are called the *output groups*.

In this way, the input group of measurements of each argument is transformed into an output group of measurement data of the measurand of the overall measurement. Combining all N resulting groups, one can find an estimate of the measurand and its accuracy. However, formally, the output groups are dependent because the transformation coefficient used to produce a given output group is determined by the averages of the input groups of its nonderiving arguments.

Statistical analysis involving dependent random quantities is generally difficult. However, ours is a specific case. It is easy to see that if we had precise values of the arguments, the transformed groups would be independent. Therefore, the dependence between the transformed groups has to do with the inaccuracy of estimates of the arguments, and this inaccuracy can be taken into account.

Assume for a moment that the indirect measurement has only two arguments, A_1 and A_2. From (5.28), if the error of the estimate of the second argument is $\delta\left(\bar{A}_2\right)$, the error of coefficient C_1 is

$$\delta(\tilde{C}_1) = \left(\frac{d\Psi_1}{dA_2}\right)_{A_2=\bar{A}_2} \delta(\bar{A}_2). \tag{5.29}$$

Because of this error, the obtained value of the measurand, $x_{1,i}$ will also have an error, which can be estimated as follows:

$$\delta\left(x_{1,i}\right) = \delta\left(\tilde{C}_1\right) f_1\left(A_{1,i}\right).$$

Dividing the above expression by $x_{1,i}$ and substituting the latter with its expression (5.27) in the right side, we obtain:

$$\frac{\delta\left(x_{1,i}\right)}{x_{1,i}} = \frac{\delta\left(\tilde{C}_1\right) f_1\left(A_{1,i}\right)}{\tilde{C}_1 f_1\left(A_{1,i}\right)} = \frac{\delta\left(\tilde{C}_1\right)}{\tilde{C}_1}.$$

Thus, we have obtained an important equation:

$$\varepsilon\left(x_{1,i}\right) = \varepsilon\left(\tilde{C}_1\right), \tag{5.30}$$

where $\varepsilon(x_{1,i})$ is the relative error of observation $x_{1,i}$ due to the inaccuracy of the transformation coefficient and $\varepsilon\big(\tilde{C}_1\big)$ is the relative errors of the transformation coefficient.

Note that the error $\varepsilon(x_{1,i})$ is the same for all members of the input group $\{x_1\}$ of the deriving argument measurements, and that it is caused by the measurement inaccuracy of the nonderiving argument A_2. Because this error is the same in all observations of the output group, the mean of the output group will have same error also. With just the data from a given indirect measurement, a point estimate of this error is impossible to find. However, having the estimate of the limits of error of the measurement of argument $\bar{A}_2$, one can estimate the inaccuracy of the corresponding transformation coefficient $\tilde{C}_1$. Then, with the latter, one can estimate the bounds on the possible changes of the mean of the output group in the case the indirect measurement is repeated. In other words, these bounds represent the limits of a possible change of the systematic error of the output group of data in the case of the repeated indirect measurement. Such an error belongs to the class of conditionally constant systematic errors. Observe that in the method of transformation, the conditionally constant systematic error in question occurs even if the measurements of the arguments have only random errors. This is because in each output group, the random errors of the means of all the nonderiving arguments are "frozen" and thus become constant.

It follows from the above discussion that accounting for the conditionally constant systematic error that arises in the method of transformation is equivalent to accounting for the dependency between the output groups of data. The estimation of the limits of this error and combining it with the random error of the measurement result is done in a usual way; it will be elaborated below.

Considering an arbitrary number of arguments now, (5.29) will take the form:

$$\delta\left(\tilde{C}_d\right) = \sum_{j \neq d} \left(\frac{\partial \Psi_d}{\partial A_j}\right)_{A_1,\ldots A_{d-1},A_{d+1},\ldots A_N = \bar{A}_1,\ldots \bar{A}_{d-1},\bar{A}_{d+1},\ldots \bar{A}_N} \cdot \delta\left(\bar{A}_j\right),$$

where all the partial derivatives are evaluated in the point $\left\{\bar{A}_j\right\},\ j \neq d$.

Generalizing the results obtained for two arguments, we can write:

$$\varepsilon\left(x_{d,i}\right) = \varepsilon\left(\tilde{C}_d\right) = \sum_{j \neq d} \frac{w_j}{\tilde{C}_d} \delta\left(\bar{A}_j\right), \tag{5.31}$$

where w_j is the influence coefficient of argument A_j and is equal to $w_j = \partial\Psi/\partial A_j$ computed in the point $\left\{\bar{A}_j\right\},\ j \neq d$. It follows from (5.31) that the confidence interval of the conditionally constant error will need to be estimated using the traditional method. Note, however, that while the traditional method here will have all the drawbacks we mentioned in the beginning of this section, these drawbacks now only apply to conditionally constant errors, while before they applied to the entire errors. Thus, while the drawbacks are the same, their effect is now reduced. With two arguments, the confidence interval for the conditionally constant error will be more accurately since in this case (5.31) has only one item.

We are now ready to estimate the measurand and its inaccuracy. All output groups of data represent "observed" values of the measurand. Therefore, as discussed later in Chap. 7, they can be combined into one large group, with its mean used as the estimate of the measurand:

$$\bar{\bar{x}} = \frac{\sum_{d=1}^{N} \sum_{i=1}^{n_d} x_{d,i}}{\sum_{d=1}^{N} n_d}. \tag{5.32}$$

Note that because in transforming input groups of argument data into output groups of the measurand data we do not use Taylor's series, in our case,

$$E[\bar{\bar{x}}] = E\big[f(A_1, \ldots, A_N)\big].$$

We can now estimate the variance and standard deviation of the mean, using (4.18) for direct measurements. But as we know, it is desirable to estimate these characteristics in relative form. Thus, we have:

$$S^2_{\Psi,rel}(\bar{\bar{x}}) = \frac{1}{\bar{\bar{x}}^2} \frac{\sum_{k=1}^{Z} (x_k - \bar{\bar{x}})^2}{Z(Z-1)}, \quad S_{\Psi,rel}(\bar{\bar{x}}) = \sqrt{S^2_{\Psi,rel}(\bar{\bar{x}})}, \tag{5.33}$$

where $Z = \sum_{d=1}^{N} n_d$ is the number of items in the combined output group, and $x_k (k = 1, \ldots, Z)$ are the items in this combined group. Knowing the estimate of the standard deviation, we can calculate the uncertainty due to random error.

Accounting for systematic error is a bit more complex. First, we need to find the confidence intervals for each output group of data. This does not present a difficulty because usually the arguments are measured by direct measurements. Then, the variance and standard deviation for each group d are estimated using formula (4.18) – again in relative form. The degree of freedom is known precisely: $\nu_d = n_d - 1$. Confidence probability α must be selected the same for every group. Then, we find quantiles $t_{d,\vartheta}$ of Student's distribution corresponding to the chosen confidence probability. If the number of measurements of each argument is the same, the quantiles will also be the same for each output group: $t_{d,\vartheta} = t_\vartheta$. From this quantile, we can find the confidence limits $\pm\theta_{d,r}$ for each output group. But to do that, we need to find standard deviation $S_{\vartheta,rel}(\bar{x}_d)$.

Using (5.31), we can calculate the variance of the measurand estimate in each group d due to the conditionally constant error. In relative form, this variance is as follows:

$$S^2_{\vartheta,rel}(\bar{x}_d) = \sum_{j \neq d} \left(\frac{w_j}{\tilde{C}_d}\right)^2 S^2_{rel}(\bar{A}_j). \tag{5.34}$$

Having found the variance estimate above, and hence the standard deviation, the corresponding confidence limit is as follows:

$$\theta_{d,rel} = t_{d,\vartheta} S_{\vartheta,rel}(\bar{x}_d). \tag{5.35}$$

If each output group had the same "frozen" (constant error), the same error would be present in the mean of the combined group. However, one cannot count on these errors to match among all the groups. Thus, it is reasonable to compute the conditionally constant error of the combined mean, θ_{rel}, as a weighted mean of the conditionally constant errors of individual output groups, with weights g_d equal to the ratio of the number of items in each group over the total size of the combined group. In other words,

$$\theta_{rel} = \sum_{d=1}^{N} g_d \theta_{d,rel} \quad g_d = n_d / Z. \tag{5.36}$$

The same weights allow us to compute the standard deviation of the conditionally constant error of the measurement result (which is the mean value of the combined output group of data):

$$S_{\vartheta,rel}\left(\bar{\bar{x}}\right) = \sum_{d=1}^{N} g_d S_{\vartheta,rel}\left(\bar{x}_d\right). \tag{5.37}$$

Now, following the procedure described in Chap. 4, we can compute the confidence interval of the measurement result. The calculations involve several steps.

First, we compute the combined standard deviation according to (4.19):

$$S_c = \sqrt{S_{\vartheta,rel}^2 + S_{\Psi,rel}^2}.$$

Next, using (4.22), we find coefficient t_c as weighted mean between t_q and t_ϑ. Coefficient t_q is found using Student's distribution with the degree of freedom $\nu = Z - 1$. Coefficient t_ϑ is obtained either from (4.21), in which case it is $t_\vartheta = \theta_{rel} / S_{\vartheta,rel}$, or, if all arguments were measured the same number of times n, from Student's distribution with $\nu = n - 1$. We should stress again that in using Student's distribution, one must select the same confidence probability in all cases.

Finally, we obtain the uncertainty of the measurement result in relative form:

$$u_{c,rel} = t_c S_c,$$

which provided the solution to the problem.

We should note that measurements of the arguments could have their own systematic errors, in most cases conditionally constant ones. They must be taken into account. To this end, for each argument, we must combine its conditionally constant and random errors. This task is accomplished using the general scheme considered in Chap. 4 and which we just used in combining errors of the overall result.

As already mentioned, our calculation procedure has a drawback: to compute the error of the transformation coefficient, we use the traditional method, which reduces somewhat the accuracy of the method although not as much as if we used the traditional method to estimate the error of the entire indirect measurement. But if measurement has only two arguments, this drawback disappear: in this case the estimation of the conditionally constant errors in each group do not require summation.

Thus, this drawback can be removed if we combined the output groups in a pairwise manner rather than all together at once. We describe this modified procedure next.

Referring to the measurement equation expression of (5.26), the calculations for the indirect measurement processing can be accomplished by a series of successive argument substitutions. Each step of this process substitutes a pair of arguments with one new argument. After $(N-2)$ steps, the original equation with N arguments will be transformed into an equivalent measurement equation having only two arguments. The processing at each step, as well as handling of the final equation, uses the same simple calculations based on the method of transformation for a measurement with two arguments.

To illustrate the main idea of this method, consider an indirect measurement with four arguments:

$$x = f_1(A_1) \bullet f_2(A_2) \bullet f_3(A_3) \bullet f_4(A_4).$$

We start by substituting the first two arguments, A_1 and A_2. To this end, we replace the corresponding terms with a new argument $B' = f_1(A_1) \bullet f_2(A_2)$. The measurement equation now becomes

$$x = B' \bullet f_3(A_3) \bullet f_4(A_4).$$

We now apply the method of transformation to the expression for B' above. Since we only have two arguments, this method is more precise due to precisely known degree of freedom of both arguments. According to this method, we use the measurement data for arguments A_1 and A_2 to obtain the data set for B', $\{B'_i\}$, $i = 1$, $(n_1 + n_2)$, and from it the estimate $\tilde{B}'$ and its standard deviation, to be used in the next step. As we mentioned earlier, the two output groups comprising $\{B'_i\}$ can be shifted against each other, but this will be taken into account when computing the variance of the combined group according to (7.10) from Chap. 7.

Continuing the substitution process, we substitute the first pair of arguments in the equation that resulted from the previous step, B' and A_3, with a new argument $B'' = B' \bullet f_3(A_3)$. Similar to the first step, we use the data set for B', along with its estimate and standard deviation (from the previous step), as well as the measurement data for A_3, to produce the set $\{B''_i\}$, $i = 1$, $(n_1 + n_2 + n_3)$ for argument B'', its estimate $\tilde{B}''$, and its standard deviation. Again, any systematic shift in subgroups of $\{B''{}_i\}$ will be taken into account by (7.10).

The measurement equation after the last step contains only two arguments:

$$x = B'' \bullet f_4(A_4).$$

Using the data set and estimate for B'' and the measurement data for A_4, we can now obtain the data set for the measurand x, $\{x_k\}$, $k = 1, \dots, \sum_{j=1}^{N} n_j$. This last set, along with the standard deviation of B'', allows us to obtain the estimate of the measurand and its uncertainty.

A detailed example of using the method of transformation is presented later in Sect. 8.6.2.

5.7 Total Uncertainty of Indirect Measurements

The preceding sections of this chapter considered multiple indirect measurements that did not have systematic errors. But the systematic errors cannot be ignored – they have to be taken into account when computing the overall inaccuracy of indirect measurements.

Systematic errors are not apparent in the process of measurements, and therefore, they must be evaluated, taking into account the possible causes of them: first, the systematic errors in the measurements of arguments. The calculations for estimating these errors are the same for the dependent and independent indirect measurements.

The relationship between the measurement errors of arguments and the error of the indirect measurement is represented by (5.13). This equation reflects the transformation of the errors in measurements of arguments into the error of an indirect measurement.

In addition to the error from the measurement errors of arguments, the indirect measurements have an additional source of error. It is the inaccuracy of the measurement equation. The next example will illustrate this error.

Suppose that we are required to measure the area of a plot of land that is depicted by a rectangle on a sketch. Here, the rectangle is the model of the object. Its area is $S_m = ab$ where a and b are the lengths of the sides of the rectangle. The discrepancies between the model and the object can in this case stem the fact that the angle between the sides will not be exactly 90°, that the opposite sides of the area will not be precisely identical, and that the lines bounding the area will not be strictly straight. Each discrepancy can be estimated quantitatively and then the error introduced by it can be calculated. It is usually obvious beforehand which source of error will be most important.

Suppose that in our example the most important source of error is that the angle between adjoining sides differs from 90° by β, as shown in Fig. 5.1. Then the area of the plot would have to be calculated according to the formula $S_t = ab\cos\beta$. Therefore the error from the threshold discrepancy in this case will be

$$S_m - S_t = ab\,(1 - \cos\beta)\,.$$

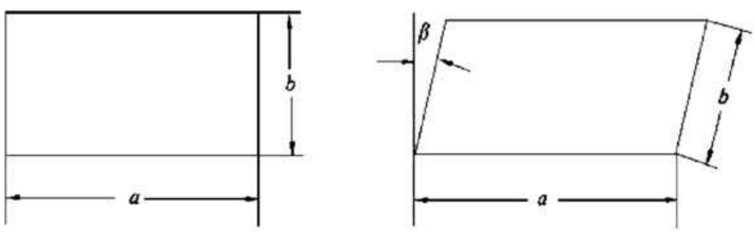


Fig. 5.1 Rectangle and parallelogram as the models of a plot of land

The admissible angle β_a must be estimated from the required accuracy in determining the area of the plot. If $\beta \geq \beta_a$, then the model must be redefined and the measured quantity must be defined differently. Correspondingly, we must use a different formula for calculating the measured area.

We should note that the inaccuracy of the measurement equation, or the threshold discrepancy between the model of an object to be studied and the object, is a methodological error and it is an absolutely constant systematic error.

The random errors of indirect measurements were analyzed previously in this chapter. Let us now begin the analysis of the systematic errors of indirect measurements.

The general approach to the problem of the estimation of systematic errors is similar to the one developed for direct measurements. Still, indirect measurements have some specifics. One difference has to do with the existence of argument influence coefficients w_j. Usually their values are calculated by substituting the estimates of arguments for their true values. In other cases, these coefficients are found from special experiments. Either way, they are obtained with some errors. These errors can be avoided if the measurement equation has the form

$$A = A_1^{l_1} A_2^{l_2} \cdots A_N^{l_N}. \tag{5.38}$$

In this case, the influence coefficients are determined by the expressions

$$\begin{aligned} w_1 &= \frac{\partial A}{\partial A_1} = l_1 A_1^{l_1 - 1} A_2^{l_2} \cdots A_N^{l_N} \\ w_1 &= \frac{\partial A}{\partial A_2} = A_1^{l_1} l_2 A_2^{l_2 - 1} \cdots A_N^{l_N} \\ &\cdots \\ w_N &= \frac{\partial A}{\partial A_N} = A_1^{l_1} A_2^{l_2} \cdots l_N A_N^{l_N - 1} \end{aligned}$$

The absolute error is determined by (5.13). We shall now transfer from the absolute error to the relative error:

$$\varepsilon = \frac{\tilde{A} - A}{A} = \frac{l_1 A_1^{l_1 - 1} A_2^{l_2} \cdots A_N^{l_N}}{A} \zeta_1 + \cdots + \frac{l_N A_1^{l_1} A_2^{l_2} \cdots A_N^{l_N - 1}}{A} \zeta_N$$

Substituting (5.38) for A, we obtain

$$\varepsilon = l_1 \frac{\zeta_1}{A_1} + l_2 \frac{\zeta_2}{A_2} + \cdots + l_N \frac{\zeta_N}{A_N}.$$

Thus, the influence coefficients for the relative errors in the measurements of the arguments are equal to the powers of the corresponding arguments: $w'_1 = l_1$, $w'_2 = l_2, \ldots, w'_N = l_N$. The coefficients $l_1, l_2 \ldots l_N$ are known exactly a priori, so that the error of influence coefficients noted above does not arise.

This result can be obtained without use of (5.13), in other words, without the use of Taylor series. Indeed, moving from (5.38) to the differentials on both sides, we obtain:

$$dA = l_1 A_1^{l_1-1} A_2^{l_2} \cdots A_N^{l_N} dA_1 + l_2 A_1^{l_1} A_2^{l_2-1} \cdots A_N^{l_N} dA_2 + \cdots + l_N A_1^{l_1} A_2^{l_2} \cdots A_N^{l_N-1} dA_N.$$

Dividing both sides by A, and replacing A with its expression (5.38) on the right side of the above equation, we get:

$$\frac{dA}{A} = l_1 \frac{dA_1}{A_1} + l_2 \frac{dA_2}{A_2} + \cdots + l_N \frac{dA_N}{A_N}.$$

Because measurement errors are small, the differentials above can be replaced by increments – measurement errors. This brings the above equation to the same expression for the combined error ε that was obtained above.

So, relative form of errors provides the uncertainty calculations with exact values of influence coefficients. This is another advantage of expressing the measurement errors in the form of relative errors.

The systematic error of the measurement of each argument consists of elementary components. As always, they can be divided into two categories: absolutely and conditionally constant errors.

Absolutely constant errors are deterministic quantities. However, we cannot find their exact values and can only estimate their limits. These limits are estimated differently in every specific case. In general, these estimations are based on the experience of the person performing the measurement. Usually, there are very few such errors and they are small. But it is necessary to keep them in mind. One example of absolutely constant errors is the error in a measurement equation considered above, such as the linearization error of the standardized characteristic of a thermocouple.

Conditionally constant errors can be computed using the first term of (5.14):

$$\vartheta = \sum_{j=1}^{N} w_j \vartheta_j,$$

where ϑ is the conditionally constant error of an indirect measurement, and ϑ_j is the conditionally constant error of estimate of jth argument. This formula can be represented in the form

$$\vartheta = \sum_{j=1}^{N} \sum_{i=1}^{k_j} w_j \vartheta_{ji}, \tag{5.39}$$

where k_j is the number of conditionally constant errors in the measurement of the jth argument.

We previously considered one difference between estimating systematic errors in the case of direct and indirect measurements, namely the existence of the influence coefficients of the arguments. But there is also another difference: In the case of dependent indirect measurements, some elementary errors in the measurements of different arguments are caused by the *same* influence quantity. When such a quantity grows, some of these errors can grow also while the rest of them go in the opposite direction.

For example, assume that two measuring instruments used in an indirect measurement have temperature errors. When the temperature changes, these errors will also change, and both of them can change either in the same direction or in opposite directions. Thus, the additional errors caused by the same influence quantity can to some degree cancel each other. In order to take advantage of such error cancellation, one must combine the additional errors caused by the same influence quantity *before* summing up the squared limits of the elementary errors. Let us consider these calculations.

For simplicity, we will consider an indirect measurement with four arguments ($N = 4$). We will further assume that the measurements of arguments 1 and 2 have additional errors caused by a change of influence quantity t, for example, temperature. Denote these additional errors ϑ_{1t} and ϑ_{2t}, respectively. They cause the resulting measurement error $\vartheta_{1,2t}$ equal to

$$\vartheta_{1,2t} = w_1\vartheta_{1t} + w_2\vartheta_{2t} \tag{5.40}$$

Taking into consideration that ϑ_{1t} and ϑ_{2t} are just two of the errors of arguments 1 and 2, and that we have four arguments altogether, (5.39) becomes as follows:

$$\vartheta = \vartheta_{1,2t} + w_1\sum_{i=1}^{k_1-1}\vartheta_{1i} + w_1\sum_{i=1}^{k_2-1}\vartheta_{2i} + w_1\sum_{i=1}^{k_3}\vartheta_{3i} + w_1\sum_{i=1}^{k_4}\vartheta_{4i} \tag{5.41}$$

As was discussed in Chap. 4, it is possible to assume all conditionally constant errors to be random quantities with a uniform distribution, and the confidence limits of the conditionally constant error of an indirect measurement θ_α can be calculated from the limits of the conditionally constant elementary errors using the same method that was discussed in Chap. 4. The main difference is that now we have to account for influence coefficients of the arguments. So, adding these coefficients to (4.3), we get

$$\theta_\alpha = k\sqrt{\theta_{1,2t}^2 + w_1^2\sum_{i=1}^{k_1-1}\theta_{1,i}^2 + w_2^2\sum_{i=1}^{k_2-1}\theta_{2,i}^2 + w_3^2\sum_{i=1}^{k_3}\theta_{3,i}^2 + w_4^2\sum_{i=1}^{k_4}\theta_{4,i}^2}, \tag{5.42}$$

and

$$\theta_{1,2t} = w_1\theta_{1t} + w_2\theta_{2t}, \tag{5.43}$$

where (5.43) is computed while preserving the signs of θ_{1t} and θ_{2t}.

The values of k are given in Sect. 4.4. In particular, for the confidence probability $\alpha = 0.95$, $k = 1.1$.

If the indirect measurement is performed under reference conditions for all instruments involved, or if no influence quantity causes additional errors in more than one instrument, then (5.42) has the form

$$\theta_\alpha = k\sqrt{\sum_{j=1}^{N} w_j^2 \theta_j^2} \tag{5.44}$$

Because all conditionally constant errors were taken to be uniformly distributed random quantities, the standard deviation of their sum can be computed as follows:

$$S_\vartheta = \frac{1}{k\sqrt{3}}\sqrt{\theta_{1,2t}^2 + w^2\sum_{i=1}^{k_1-1}\theta_{1,i}^2 + \cdots + w_n^2\sum_{i=1}^{N}\theta_{N,i}^2} = \frac{\theta_\alpha}{k\sqrt{3}}.$$

Now let us return to the absolutely constant errors. Adding up their limits, we obtain the overall limits H of the absolutely constant error of the result of an indirect measurement:

$$H = H_e + \sum_{j=1}^{N} w_j H_j$$

where H_e is the limit of an error of the measurement equation; H_j is the limit of the absolutely constant error of the measurement of jth argument.

Thus, we have the estimate of the variance of conditionally constant errors S_ϑ^2 and the limits of the absolutely constant error H. We also have the estimate of the variance of the random error $S_{\bar{x}}^2$. So, we can now obtain the total uncertainty of indirect measurement result. These calculations are exactly the same as those used for the uncertainty calculation in Chap. 4 for direct measurements. Therefore, in the same way, we can now calculate the uncertainty of indirect measurements. The resulting formulas are repeated below.

The combined standard deviation S_c can be calculated using (4.19):

$$S_c = \sqrt{S_\vartheta^2 + S_{\bar{x}}^2}. \tag{5.45}$$

The combined uncertainty can be found from (4.20):

$$u_c = t_c S_c, \tag{5.46}$$

and the coefficient t_c is calculated by (4.22):

$$t_c = \frac{\theta_\alpha + t_q S_{\bar{x}}}{S_\vartheta + S_{\bar{x}}} = t_\vartheta \frac{S_\vartheta}{S_\vartheta + S_{\bar{x}}} + t_q \frac{S_{\bar{x}}}{S_\vartheta + S_{\bar{x}}}. \tag{5.47}$$

Because $S_\vartheta = \frac{\theta_\alpha}{k\sqrt{3}}$, t_ϑ depends only of confidence probability α. If $\alpha = 0.95$, $k_{0.95} = 1.1$ and $t_\vartheta = 1.90$.

Taking into account the limit of the absolutely constant error, we obtain the total uncertainty u_t of the measurement result:

$$u_t = H + u_c. \tag{5.48}$$

5.8 Accuracy of Single Indirect Measurements

Single indirect measurements are very important in practice but unlike direct measurements, they cannot be viewed as the base form of multiple indirect measurements. As we discussed, this is due to the fact that in multiple indirect measurements, it is the arguments that are measured multiple times rather than the indirect measurement being repeated.

Among examples of single indirect measurements, we can list measurement of the area of a plot of land, measurement of wattage dissipated by a resistor under high-frequency current, and measurement of temperature using separately calibrated thermocouple and millivoltmeter.

In single indirect measurements, the estimate of the measurand is obtained by putting the estimates of all the arguments into the measurement equation. The estimates of the arguments and their inaccuracy are typically obtained using direct measurements. We have described the methods to accomplish these tasks in Chap. 4.

The estimation of inaccuracy of single indirect measurements is in principle analogous to that of direct measurements; the only difference is that in measurements under reference conditions, the inaccuracy of direct measurements is determined by the intrinsic error of a single measuring instrument while in indirect measurements, of several instruments. Therefore, inaccuracy of indirect measurements involves summation of errors even under reference conditions whereas in direct measurements, this is only needed under rated conditions. The summation methods themselves remain the same. The fact that errors of argument measurements must be viewed as elementary errors (even though each argument has its own elementary errors) and that the number of elementary errors in the case of indirect measurements is typically greater is not principally significant. However, the calculation formulas take a different form because the meaning of influence coefficients changes. Consequently, we rewrite these formulas below.

1. *Measurements under reference conditions for all instruments involved.* The inaccuracy of measurements of the arguments is expressed in the form of limits of error Δ_j for each argument $A_j (j = 1, \ldots, N)$. These limits are transformed into the limits of elementary error of indirect measurement θ_j as follows:

$$\theta_j = w_j \Delta_j,$$

where $w_j = \frac{\partial f}{\partial A_j}$ is the influence coefficient of argument A_j computed at the point with coordinates $\left(\tilde{A}_j\right), j = 1, \ldots, N$.

As explained in Chap. 4, we can take a uniform distribution for the model of elementary errors with given limits. Further, in Sect. 4.4, we proposed and analyzed a method for summation of the limits of uniform distributions, and we applied this method for summation of the elementary errors of single direct measurements under rated conditions in Sect. 4.7. Thus, we will utilize the recommendations formulated in Sect. 4.7, taking into account that the measurement errors of the arguments, multiplied by the corresponding argument influence coefficients, become elementary errors of the indirect measurement. Accordingly, (4.3), which expresses the uncertainty of a single measurement, becomes

$$u_\alpha = k\sqrt{\sum_{j=1}^{N} w_j^2 \Delta_j^2} = k\sqrt{\sum_{j=1}^{N} \theta_j^2}. \tag{5.49}$$

One must remember that the argument influence coefficients obtained from calculations have certain inaccuracy. This inaccuracy can often be avoided by representing the errors in relative form (see Sect. 5.7). Thus, expressing errors in relative form is preferable.

From the discussion in Sect. 4.7, it follows that with confidence probability $\alpha = 0.95$, (5.49) can be used with any number of component errors, and with the same value of $k_{0.95} = 1.1$. With $\alpha = 0.99$, the calculations depend on the number of components and are the same as with direct measurements under rated conditions (see Sect. 4.7).

2. *Measurements under rated conditions.* When some of the instruments are used under rated conditions, one must account for additional errors besides the intrinsic errors. There are two ways to combine them. One method involves estimating the measurement uncertainty of each argument and then combining them. The other combines elementary measurement errors of all the arguments. The latter method appears preferable because all errors being combined become homogeneous in a sense that they all are specified by their limits. Therefore, they can be combined according to the same recommendations that were described in Sect. 4.7 for direct measurements. The one peculiarity arising in indirect measurements is due to the fact that additional errors in different instruments can be caused by the same influence quantity and therefore can be mutually dependent. Accounting for this dependency is considered in Sect. 5.7.

5.9 Accuracy of a Single Measurement with a Chain of Instruments

Single measurements are often performed using several measuring instruments connected in a chain. A chain of serially connected instruments is also commonly called a measurement system. When using an instrument chain, the measurement result is given by the indication of the last instrument.

A simple example of such measurements is the measurement of temperature with thermocouple and millivoltmeter. The thermocouple produces for each temperature T_x the corresponding electromotive force (EMF) U, and the voltmeter measures this EMF. The measurement equation is

$$T_x = KU,$$

where K is the thermopower of the thermocouple. Because of the nonlinear dependency of the EMF of the thermocouple on the temperature, the thermopower itself depends on the temperature.

Thermocouple characteristics, which specify the relationship between the temperature and the EMF, are standardized, and knowing the type of thermocouple, one can find the temperature T_x, as well as the thermopower K, corresponding to a given indication of the voltmeter U. If the voltmeter had been graduated in the units of temperature, its indications will produce temperature as the result of the measurement, and so the measurement must be considered a direct measurement. If the voltmeter is graduated in volts, then it becomes an indirect measurement. Let us calculate the inaccuracy of this measurement.

The inaccuracy of this measurement is calculated based on the known limits of admissible errors of the voltmeter, Θ_1, and thermocouple, Δ_2, in the given point of its characteristic. Expanding the measurement equation into Taylor series produces:

$$t_x = Tx + (w_U \Delta_1 + w_K \Delta_2) + (w_U \Delta_1 + w_K \Delta_2)^2 + \cdots,$$

where w_U and w_K are argument influence coefficients computed as partial derivatives of the measurement equation in point (u, k): $w_U = k$ and $w_K = u$.

Discarding as usual the terms in the second and higher degrees and consider that Δ_t is absolutely constant error, we obtain:

$$u_t = t_x - T_x = k\Theta_1 + u\Delta_2,$$

where u_t can be considered as the limit of error of the measurement result if $u\Delta_2 > w_K \Theta_1$.

In a general case the chain can have not two but N instruments. The measurement equation, however, will retain its structure. Therefore, the measurement error will still be represented by the sum of the limits of error of the instruments in the chain multiplied by the corresponding argument influence coefficients.

If the instruments were utilized under reference conditions, one can combine the component errors in the exact same way as described in Sect. 5.8.

An example of a serial connection of several instruments is described in detail in Chap. 8 (Sect. 8.2), where we consider a measurement of voltage with a potentiometer, a voltage divider, and a standard cell.

5.10 Application of the Monte Carlo Method

The Monte Carlo method is a numerical method of obtaining a composition of independent random quantities with known distribution functions. In the old days of manual computations this method used to be too laborious to be used in measurements, but thanks to modern computers, it can now be employed widely. The wide adoption of the Monte Carlo method should be facilitated by recommendation [13], which is to be finalized in the near future.

The essence of the Monte Carlo method can be explained as follows. For simplicity, let us consider random quantity Z related with a known dependency f with only two independent random quantities X and Y:

$$Z = f(X, Y).$$

Using a random numbers generator and known distribution functions, we obtain a series of realizations of X and Y. According to [13], one should have 10^6 realizations of each random quantity. Randomly chosen pairs of realizations x and y are input into dependency f to obtain a realization of Z. Once a pair of input realization is utilized, it is excluded from further calculations (i.e., we utilize sampling without replacement), so that each pair produces one realization of Z. A large number of realizations of Z allow one to construct the distribution function of this random quantity and compute its mathematical expectation, variance, and the confidence interval for the true value of Z.

Thus, the Monte Carlo method gives a full solution to the problems of experimental data processing for independent indirect measurements. However, as we showed earlier in this chapter, much simpler and complete solutions to these problems exist for independent indirect measurements. Nonetheless, the Monte Carlo method has its place in that it can be used to verify the accuracy of simple methods or as an alternative method in doubtful cases. Thus, we will now consider in more detail the application of the Monte Carlo method to practical measurements.

The key problem in applying the Monte Carlo method in measurements is that the distribution functions of the measurement results of arguments are unknown. The standard draft [13] lists a number of predefined distribution functions but does not say how one can choose one of these functions based on the experimental data available. The fundamental question of how the discrepancy between the experimental data and the chosen distribution function would affect the result produced by the Monte Carlo method remains open.

A possible way to address this issue is to use analytical approximation of the distribution functions derived from the available experimental data in place of the actual distribution functions. The overall method then involves the following four stages.

1. Use the measurement results of each argument to approximate its probability distribution function. Any approximation method could be used; for example, one could build a histogram of the measurement results to approximate the density

function and then obtain the cumulative distribution function from it by a numerical integration. At the end of this stage, we have an approximation of the distribution function of every argument.

2. Using the simulation method, use a generator of uniformly distributed random numbers to obtain a given number of "virtual measurements" of each argument (or errors of these measurements, depending on how the sample data are represented), distributed according to the corresponding analytical distribution function. This is done as follows [35]. Take, for example, argument X and assume we want to have $n = 1{,}000$ virtual observations. To obtain these observations, first obtain 1,000 random numbers from the generator and normalize them (by dividing by the maximum possible number) so that they all lie within interval [0, 1]. Next, treat them as probabilities and find the corresponding quantiles of the probability distribution function of X, using the approximation of this probability distribution found in stage 1. These quantiles can be used as realizations of X. Repeat this procedure for Y and all the other arguments. The number of realizations of each argument must be the same.
3. Substitute into the measurement equation one realization of each argument. Each set of these argument realizations will produce one realization of the measurand Z. We should stress again that every argument realization is only used once; this ensures that all generated realizations of Z will be independent.
4. At the end of stage 3, n realizations of Z will have been produced. Using these data, build a histogram of the realizations of the measurand. With a large n, the obtained histogram will closely match the experimental distribution function of Z and will allow one to find all the necessary parameters of the measurement: the estimate of the measurand and its inaccuracy.

To ensure the stability of the obtained results, it is advisable to repeat the above calculations 2–3 times.[3] If these repeated simulations indicate unstable results, one should increase the number of virtual realizations of the arguments. Note that one cannot simply combine output realizations obtained from repeated simulations because these realizations will be dependent (see "Monte Carlo Statistical Methods" by Christian P. Robert, George Casella).

The Monte Carlo method in its essence involves an artificial increase in the amount of experimental data. But it is not the only possible method to achieve this effect. Another possibility is the bootstrapping method [23], which we outlined in Chap. 3 (Sect. 3.8). An advantage of the latter is that it produces independent samples and hence the results obtained from these samples can be combined.

An even simpler method is possible if the measurement is automated. In this case, one can obtain sufficiently large samples from the experiment directly so that random combinations of the results of measurement of the arguments would produce a stable distribution of the estimates of the measurand (again, the combinations of

[3] In doing so, one must be careful to avoid a mistake of simply rerunning the random number generator from scratch using the same seed: this would produce the exact same sequence of numbers every time and will provide no indication of the stability of the results.

argument realizations should be utilized without replacement to avoid the dependencies between the obtained estimates of the measurand). Unfortunately, this approach is feasible not in all measurement fields.

Our final note regarding the Monte Carlo method is that approximations of distribution functions with analytical formulas from histograms, which are typically constructed from a small number of realizations, cannot be accurate. Moreover, these approximations are often obtained subjectively to a large degree, such as the case with the recommendations from [13]. The inaccuracy of this step limits the effectiveness of the Monte Carlo method. At the same time, the traditional method and the method of transformation solve the problem of constructing confidence intervals for the independent indirect measurements and the method of reduction solves it for dependent indirect measurements. Therefore, the more complex Monte Carlo method is not necessary in practice for these purposes. This concern does not apply, however, to the one particular metrological application of the Monte Carlo method mentioned earlier, namely, for verifying the accuracy of methods for estimating the uncertainty of indirect measurements that are used in practice. The Monte Carlo method is valuable in this application because it allows one to investigate theoretical scenarios with precisely specified distribution functions. It may also be possible that such investigations would lead to discovery of new approaches, which would be simpler than the Monte Carlo method and more accurate than existing practical methods.

Chapter 6
Combined and Simultaneous Measurements

6.1 General Remarks About the Method of Least Squares

Combined and simultaneous measurements, as pointed out in Chap. 1, are measurements performed to find values of several quantities related by a known equation. In either case, a measurement experiment involves multiple measurements, with each individual measurement producing one equation instance. Typically, the number of measurements is such that there are more equations than the unknowns (the parameters and measurands). Because of measurement errors, it is impossible to find values of the unknowns such that all equations would be satisfied simultaneously. Under these conditions, the estimated values of the unknowns usually are found with the help of the *method of least squares*.

The method of least squares is a widely employed computational technique that makes it possible to handle the inconsistency of experimental data. This method is easily implemented with the help of computers, and good least-squares software is available.

There is extensive literature on the method of least squares, and it has been well studied. It is known that the estimates obtained with this method satisfy the requirements for estimates from Sect. 3.2 only if all the errors in the measurements are random and normally distributed. Nevertheless, the method of least squares is widely employed, because it is simple and in general, the biasness of the estimates obtained is usually not significant even when the above condition does not hold. Moreover, in measurement practice, the least-squares method is also used to reduce the systematic errors if the measurement experiment can be organized in such a way that different measurements of the same quantities have different systematic errors.

An alternative to the least-squares method is the method of minimizing the sum of absolute deviations. This method is even more intuitive than the method of the least squares although it involves more complex calculations. While the advent of computers has made the complexity of calculations irrelevant, it is still seldom used.

An example of simultaneous measurements is finding the parameters of the equation that expresses the temperature dependence of an accurate measuring resistor:

$$R = R_{20} + a(t - 20) + b(t - 20)^2,$$

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*, DOI 10.1007/978-1-4419-1456-9_6,

where R is the resistance of the resister, t is its temperature, R_{20} is the resistance of the resistor at $t = 20°C$, and a and b are the temperature coefficients. By measuring simultaneously R and t and by varying the temperature, we obtain several equations, from which it is necessary to find R_{20} and the temperature coefficients. When the number of measurements exceeds 3, we cannot find an unambiguous solution, and the least-squares method can be used to find the estimates of the parameters.

Because both combined and simultaneous measurements utilize the method of least squares, and the technique is exactly the same in both cases, for brevity, we will use the term "combined measurements" in this chapter to refer to both these types of measurements. We shall now discuss the method of least squares because of its importance to combined measurements and because understanding its basic ideas is necessary to use this method properly.

We can write the basic measurement equation of the combined measurement in the general form

$$F(A, B, C, \ldots, x, y, z, \ldots) = l, \tag{6.1}$$

where x, y, z, and l are directly measured quantities, and A, B, and C are the unknowns to be determined.

Substituting the experimentally obtained numerical values of x_i, y_i, z_i, and l_i into (6.1), we obtain a series of equations of the form

$$F(A, B, C, \ldots, x_i, y_i, z_i) = l_i, \tag{6.2}$$

which contain only the unknown quantities A, B, and C to be estimated and the numerical values of the measured quantities. The quantities sought are found by solving the obtained equations simultaneously.

An example of a combined measurement is finding the capacitances of two capacitors from the measurements of the capacitance of each one of them separately, as well as when the capacitors are connected in parallel and in series. This method for measuring the capacitances of the capacitors could be chosen to reduce somewhat the systematic error of the measurement, which is different at different points of the measurement range – reducing the random component of the error could be accomplished by simply measuring each capacitance multiple times.

Each measurement is performed with one observation, but ultimately, we shall have four equations for the two unknown capacitances C_1 and C_2:

$$C_1 = x_1, C_2 = x_2, C_1 + C_2 = x_3, \frac{C_1C_2}{C_1 + C_2} = x_4.$$

Substituting into these equations the experimentally found values of x_i, we obtain a system of equations analogous to (6.2).

As we have already pointed out, the number of equations in the system (6.2) is greater than the number of unknowns, and because of measurement errors, it is impossible to find values of the unknowns such that all equations would be satisfied simultaneously. For this reason, (6.2), in contrast to normal mathematical equations, is said to be *conditional equation*. Because of the inaccuracy of measurements, when some estimates of the unknowns, $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$, are substituted into the conditional equations (6.2), we do not obtain exact equalities:

$$F\left(\tilde{A}, \tilde{B}, \tilde{C}, \ldots\right) - l_i = r_i \neq 0.$$

The quantities r_i are called *residuals*. The values of the unknowns that minimize the sum of the squares of the residuals are generally recognized as the solution of the conditional equation. This proposition was first published by Legendre and is called Legendre's principle. He further proposed a method of finding the solution according to this principle; this method is now called the method of least squares.

6.2 Measurements with Linear Equally Accurate Conditional Equations

We will first consider the case when each conditional equation is obtained under the same conditions and either with the same instruments or the instruments of the same accuracy. Thus, each equation can be viewed as equally accurate and be given equal consideration in the calculation procedure.

To simplify the presentation, we shall consider the case of three unknowns. Let the system of conditional equations have the form

$$Ax_i + By_i + Cz_i = l_i (i = 1, \ldots, n, n > 3), \tag{6.3}$$

where A, B, and C are the unknowns to be estimated, and x_i, y_i, z_i, and l_i are the results of the ith series of measurements and known coefficients.

In the general case, the number of unknowns $m < n$; if $m = n$, then the system of conditional equations can be solved uniquely, although the obtained results are burdened with errors.

If some estimates of the unknowns $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are substituted into (6.3), then we obtain the residuals

$$r_i = \tilde{A}x_i + \tilde{B}y_i + \tilde{C}z_i - l_i.$$

Because all equations are given equal consideration, we shall find estimates of A, B, and C from the condition

$$Q = \sum_{i=1}^{n} r_i^2 = \min.$$

To do so, we consider the estimates to be chosen as variables and find the values of these estimates that minimize Q in a standard way using derivatives:

$$\frac{\partial Q}{\partial \tilde{A}} = \frac{\partial Q}{\partial \tilde{B}} = \frac{\partial Q}{\partial \tilde{C}} = 0.$$

We shall find these particular derivatives and equate them to 0:

$$\frac{\partial Q}{\partial \tilde{A}} = 2\sum_{i=1}^{n} \left(\tilde{A}x_i + \tilde{B}y_i + \tilde{C}z_i - l_i\right) x_i = 0,$$
$$\frac{\partial Q}{\partial \tilde{B}} = 2\sum_{i=1}^{n} \left(\tilde{A}x_i + \tilde{B}y_i + \tilde{C}z_i - l_i\right) y_i = 0,$$
$$\frac{\partial Q}{\partial \tilde{C}} = 2\sum_{i=1}^{n} \left(\tilde{A}x_i + \tilde{B}y_i + \tilde{C}z_i - l_i\right) z_i = 0.$$

From here, we obtain a system of so-called normal equations:

$$\tilde{A}\sum_{i=1}^{n} x_i^2 + \tilde{B}\sum_{i=1}^{n} x_i y_i + \tilde{C}\sum_{i=1}^{n} x_i z_i = \sum_{i=1}^{n} x_i l_i,$$
$$\tilde{A}\sum_{i=1}^{n} y_i x_i + \tilde{B}\sum_{i=1}^{n} y_i^2 + \tilde{C}\sum_{i=1}^{n} y_i z_i = \sum_{i=1}^{n} y_i l_i,$$
$$\tilde{A}\sum_{i=1}^{n} z_i x_i + \tilde{B}\sum_{i=1}^{n} z_i y_i + \tilde{C}\sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} z_i l_i.$$

The normal equations are often written using Gauss's notation:

$$\sum_{i=1}^{n} x_i^2 = [xx], \qquad \sum_{i=1}^{n} x_i y_i = [xy], \qquad \text{and so on.}$$

It is obvious that

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i \qquad \text{and therefore} \qquad [xy] = [yx].$$

In Gauss's notation, the normal equations assume the simpler form

$$\begin{aligned} [xx]\,\tilde{A} + [xy]\,\tilde{B} + [xz]\,\tilde{C} &= [xl]\,, \\ [xy]\,\tilde{A} + [yy]\,\tilde{B} + [yz]\,\tilde{C} &= [yl]\,, \\ [xz]\,\tilde{A} + [yz]\,\tilde{B} + [zz]\,\tilde{C} &= [zl]\,. \end{aligned} \tag{6.4}$$

We call attention to two obvious but important properties of the matrix of coefficients of the unknowns in the system of equations (6.4):

1. The matrix of these coefficients is symmetric relative to the main diagonal.
2. All elements on the main diagonal are positive.

These properties are general. They do not depend on the number of unknowns, but in this example, they are shown in application to the case with three unknowns.

The number of normal equations is equal to the number of unknowns, and solving these equations by known methods we obtain estimates of the measured quantities. The solution can be written most compactly with the help of the determinants:

$$\tilde{A} = \frac{D_x}{D}, \qquad \tilde{B} = \frac{D_y}{D}, \qquad \tilde{C} = \frac{D_z}{D}, \tag{6.5}$$

where

$$D = \begin{vmatrix} [xx] & [xy] & [xz] \\ [yx] & [yy] & [yz] \\ [zx] & [zy] & [zz] \end{vmatrix}.$$

and the determinants D_x, D_y, and D_z are obtained from the principal determinant D by replacing, respectively, the first, second, and third columns with the column of free terms. For example, the determinant D_x is obtained as:

$$D_x = \begin{vmatrix} [xl] & [xy] & [xz] \\ [yl] & [yy] & [yz] \\ [zl] & [zy] & [zz] \end{vmatrix}.$$

Now we must estimate the errors of the obtained results. We can do it as follows. Each conditional equation has its own residual. The entire set of these residuals, similar to the errors of repeated direct measurements, can be characterized by its own variance. This variance can then serve as an indication of the accuracy of the obtained results.

The estimate of the above variance is calculated from the formula

$$S^2 = \frac{\sum_{i=1}^{n} r_i^2}{n - m}, \tag{6.6}$$

where r_i is the residual of conditional equation i, n is the number of conditional equations, and m is the number of unknowns. Then the estimates of the variances of the values found for the unknowns can be calculated using the formulas

$$S^2(\tilde{A}) = \frac{D_{11}}{D} S^2, \qquad S^2(\tilde{B}) = \frac{D_{22}}{D} S^2, \qquad S^2(\tilde{C}) = \frac{D_{33}}{D} S^2, \tag{6.7}$$

where D_{11}, D_{22}, and D_{33} are the algebraic complements of the elements $[xx]$, $[yy]$, and $[zz]$ of the determinant D, respectively (they are obtained by removing from the matrix of the determinant D the column and row whose intersection is the given element).

The confidence intervals for the true values of the measured quantities are constructed in a standard way, based on Student's distribution. In this case, the degree of freedom for all measured quantities is equal to $\nu = n - m$.

Sometimes unknowns are related with a strict known dependency. For example, in measuring the angles of a triangle, we know that their sum is equal to 180°. Such a dependency is called a *constraint*. If we have n conditional equations, m unknowns, and k constraints, and $n > m - k$ and $m > k$, then k unknowns can be eliminated from the conditional equations by expressing these unknowns by the remaining unknowns. Next, using the method of least square, we find the estimates of the values of $m - k$ unknowns and the estimates of their standard deviations. The degree of freedom in this case will be $\nu = n - (m - k)$. We obtain the remaining k unknowns using the constraint equations.

To find the standard deviations of these remaining unknowns, strictly speaking, one must perform another cycle of calculations with the conditional equations, in which the k previously excluded unknowns are retained and the other unknowns are excluded. However, this is rarely (if ever) done, because usually a specific problem at hand allows for a simpler method. We will see this in an example in Sect. 6.5.

6.3 Measurements with Linear Unequally Accurate Conditional Equations

In Sect. 6.2, we studied the case in which all conditional equations could be assumed to be equally accurate and thus were given equal weight in the calculations. In practice, there can be cases in which the conditional equations have different accuracy, which usually happens if equations reflecting the measurements are performed under different conditions. For instance, some measurements might be performed at one temperature while others at a different temperature, leading to different additional errors.

For unequally accurate conditional equations, the estimates of the unknowns A, B, C, ... are obtained by minimizing the expression

$$Q = \sum_{i=1}^{n} g_i r_i^2,$$

where g_i is the weight of the ith conditional equation.

The immediate question then arises: how to assign weights to the conditional equations. Currently, the specialists conducting the measurement assign these weights from their personal experience. Obviously, such an approach is objectionable because of its subjectivity. It would be desirable to have a systematic solution using objective indications of the accuracy of measurements.

One could in principle imagine such an objective method along the following lines. If we view the residual of each conditional equation as its error, we could use the variance of the residual as the indication of its accuracy. Let us refer to the variance of the residual of a conditional equation as the variance of the conditional equation for short.

Pretend for a moment that the variances σ_i^2 of the conditional equations are known. Then the weights of these equations could be obtained from the conditions:

$$\sum\nolimits_{i=1}^{n} g_i = 1,$$
$$g_1 : g_2 : \cdots : g_n = \frac{1}{\sigma_1^2} : \frac{1}{\sigma_2^2} : \cdots : \frac{1}{\sigma_n^2}.$$

(The notation in the second line means that the pair-wise ratios of the weights should be equal to the ratios of the reverses of the corresponding variances.) Thus, the weights are

$$g_i = \frac{1/\sigma_i^2}{\sum\limits_{i=1}^{n} 1/\sigma_i^2}.$$

Unfortunately, the variances of the conditional equations are unknown. One can resolve this situation when there are a large number of conditional equations. In this case, one can often divide them into groups of equations with equal accuracy. Assume that each such group has more equations than there are unknowns. Then, for each group in isolation, one can obtain the estimate of the variance of their residuals as we did in Sect. 6.2 [see formula (6.6)]. Note that, in applying (6.6), the number of unknowns remains the same as in the overall system of equations and the number of conditional equations n is the number of equations in the group. Once the variance of the residuals in a group is found, this variance is assigned to all equations in the group.

We now assume that the weights are known. The introduction of weights is equivalent to multiplying the conditional equations by $\sqrt{g_i}$. Further, the cofactors g_i will appear in the coefficients of the unknowns in the normal equations. For example, the first equation of the system of normal equations (6.4) will assume the form:

$$[gxx]\,\tilde{A} + [gxy]\,\tilde{B} + [gxz]\,\tilde{C} + [gxl] = 0,$$

where each coefficient in the above equation is a sum of terms of the form

$$[gxy] = g_1 x_1\, y_1 + g_2 x_2\, y_2 + \cdots + g_n x_n\, y_n.$$

The remaining equations in the system (6.4) will change analogously. After these transformations, the further solution of the problem proceeds in the manner described in Sect. 8.2, and finally we obtain estimates of the measured quantities and their standard deviations.

6.4 Linearization of Nonlinear Conditional Equations

For several fundamental reasons, the method of least squares has been developed only for linear conditional equations. Therefore, the cases with nonlinear conditional equations require transformation of the conditional equations into a linear form.

The general method for doing this task is based on the assumption that the incompatibility of the conditional equations is small; i.e., their residuals are small. Then, taking from the system of conditional equations as many equations as there are unknowns and solving them, we find the initial estimates of the unknowns A_0, B_0, C_0. Next, assuming that

$$A = A_0 + a,\ B = B_0 + b,\ C = C_0 + c,$$

we substitute these expressions into the conditional equations. Let

$$F(A_0 + a,\ B_0 + b,\ C_0 + c) = l_i$$

be the resulting conditional equations. We expand these equations in Taylor series and, retaining only terms with the first powers of the corrections a, b, and c, obtain

$$F\,(A_0, B_0, C_0) - l_i + \left(\frac{\partial F}{\partial A}\right)_{(A_0, B_0, C_0)} \times a + \left(\frac{\partial F}{\partial B}\right)_{(A_0, B_0, C_0)} \times b + \left(\frac{\partial F}{\partial C}\right)_{(A_0, B_0, C_0)} \times c = 0.$$

In the above equation, the partial derivatives are found at point (A_0, B_0, C_0): we differentiate the functions $F(A, B, C)$ with respect to A, B, and C, respectively, and substitute A_0, B_0, and C_0 into the obtained formulas to find their numerical values. In addition,

$$F\,(A_0, B_0, C_0) - l_i = r_i \neq 0.$$

Thus, we have a system of linear conditional equations for a, b, and c. We can now use the method of least squares to find their estimates, $\tilde{a}$, $\tilde{b}$, and $\tilde{c}$, and standard deviations. Then

$$\tilde{A} = A_0 + \tilde{a}, \quad \tilde{B} = B_0 + \tilde{b}, \quad \tilde{C} = C_0 + \tilde{c}.$$

As A_0, B_0, and C_0 are nonrandom quantities, $S^2\left(\tilde{A}\right) = S^2\left(\tilde{a}\right)$, $S^2\left(\tilde{B}\right) = S^2\left(\tilde{b}\right)$, and $S^2\left(\tilde{C}\right) = S^2\left(\tilde{c}\right)$. In principle, once $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ have been obtained, we can repeat the above calculations with these values, instead of A_0, B_0, and C_0, as the current estimates to construct the second approximation, and so on.

In addition to the above method of linearization of the conditional equations, one can also use the *method of substitutions*. If, for example, a conditional equation has the form

$$y_i = x_i \sin A + z_i\, e^{-2B},$$

where x, y, and z are directly measured quantities, and A and B must be determined, then the substitution

$$U = \sin A,\ E = e^{-2B}$$

can be made. Then we obtain the linear conditional equation

$$y_i = x_i U + z_i\, E.$$

The solution of these equations gives $\tilde{U}$ and $\tilde{E}$ and the estimates of their variances, which can then be used to find the required quantities A and B.

The method of substitutions is convenient, but it is not always applicable. In principle, one can imagine one other general method for solving a system of equations when the number of equations is greater than the number of unknowns. This method is as follows.

Take from the available conditional equations a group of equations such that their number is equal to the number of unknowns. Such a group gives a definitive value for each unknown. Next, replacing in turn the equations in the group by each of the other equations that were not in the group, we obtain other values of the same unknowns. For each possible combination, the values of the unknowns can be found. As a result of such calculations, we produce a set of values for each unknown, which could be regarded as the group of observations obtained with direct measurements.

This method seems intuitive and attractive, but, unfortunately, it is incorrect. The problem is that the sets of values obtained for the unknowns are not independent. This presents difficulties in estimating the variances of the obtained estimates for the unknowns.

6.5 Examples of the Application of the Method of Least Squares

The examples below are presented to demonstrate the computational technique as well as the physical meaning of the method. For this reason, these examples were chosen so that the calculations would be as simple as possible. The initial data for the examples are taken from [37]. Note that, strictly speaking, the examples presented here are not combined or simultaneous measurements because all the parameters in the equations involved are known. These are rather examples where one uses the least square method to reconcile multiple measurements of several measurands whose values are constrained by known dependencies.

Example 6.1. Determine the angles of a trihedral prism. Each angle is measured three times. The measurements of all angles are equally accurate. The results of all single measurements are as follows:

$$\begin{aligned}
x_1 &= 89°55', & y_1 &= 45°5', & z_1 &= 44°57',\\
x_2 &= 89°59', & y_2 &= 45°6', & z_1 &= 44°55',\\
x_3 &= 89°57', & y_3 &= 45°5', & z_3 &= 44°58',
\end{aligned}$$

We have three unknowns – the angles – and each measurement produces one conditional equation, relating one of the unknowns to its measurand. Thus, denoting the unknown angles as A, B, and C, we have the system of nine conditional equations:

$$\begin{aligned} A &= 89°55', \quad B = 45°5', \quad C = 44°57', \\ A &= 89°59', \quad B = 45°6', \quad C = 44°55', \\ A &= 89°57', \quad B = 45°5', \quad C = 44°58'. \end{aligned}$$

If each angle is found as the arithmetic mean of the corresponding observations, then we obtain

$$A_0 = 89°57', \qquad B_0 = 45°5.33', \qquad C_0 = 44°56.67',$$

The sum of the angles of the triangle must satisfy the constraint $A + B + C = 180°$. However, we obtain $A_0 + B_0 + C_0 = 179°59'$. This discrepancy is the result of measurement errors. The values of the estimates must be changed so that the constraint is satisfied.

We now proceed to the solution of the problem. To simplify the calculations, we shall assume that

$$A = A_0 + a, \qquad B = B_0 + b, \qquad C = C_0 + c,$$

and we shall find the values of the corrections a, b, and c.

The system of conditional equations transforms into the following system:

$$\begin{aligned} a &= -2', \quad b = -0.33', \quad c = +0.33', \\ a &= +2' \quad b = +0.67', \quad c = -1.67', \\ a &= 0', \quad b = -0.33', \quad c = +1.33'. \end{aligned}$$

The constraint equation will assume the form

$$A_0 + a + B_0 + b + C_0 + c = 180°.$$

Therefore

$$a + b + c = 180° - 179° \, 59' = 1'.$$

We exclude c from the conditional equations using the relation

$$c = 1' - a - b,$$

We thus obtain the following system of conditional equations:

$$1 \times a + 0 \times b = -2', \quad 0 \times a + 1 \times b = -0.33', \quad 1 \times a + 1 \times b = +0.67',$$

$$1 \times a + 0 \times b = +2', \quad 0 \times a + 1 \times b = +0.67', \quad 1 \times a + 1 \times b = +2.67',$$
$$1 \times a + 0 \times b = 0', \quad 0 \times a + 1 \times b = -0.33', \quad 1 \times a + 1 \times b = -0.33'.$$

We now construct the system of normal equations. Its general form will be

$$[xx]\,a + [xy]b = [xl],$$
$$[xy]\,a + [yy]b = [yl].$$

Here, we obtain:

$$\begin{aligned}
[xx] &= 1 + 1 + 1 + 1 + 1 + 1 = 6,\\
[xy] &= 1 + 1 + 1 = 3,\\
[yy] &= 1 + 1 + 1 + 1 + 1 + 1 = 6,\\
[xl] &= -2' + 2' + 0.67' + 2.67' - 0.33' = +3',\\
[yl] &= -0.33' + 0.67' - 0.33' + 0.67' + 2.67' - 0.33' = +3'.
\end{aligned}$$

Therefore, the normal equations will assume the form

$$6a + 3b = 3', \quad 3a + 6b = 3'.$$

In accordance with the relations (6.5), we calculate

$$D = \begin{vmatrix} 6 & 3 \\ 3 & 6 \end{vmatrix} = 36 - 9 = 27,$$

$$D_a = \begin{vmatrix} 3' & 3 \\ 3' & 6 \end{vmatrix} = 18' - 9' = 9'.$$

$$D_b = \begin{vmatrix} 6 & 3' \\ 3 & 3' \end{vmatrix} = 18' - 9' = 9',$$

and we find

$$\tilde{a} = \tilde{b} = 9'/27 = 0.33'.$$

Therefore, $\tilde{c} = 0.33'$ also.

Substituting the obtained estimates into the conditional equations, we calculate the residuals:

$$\begin{array}{lll}
r_1 = 2.33' & r_4 = 0.67' & r_7 = 0\\
r_2 = 1.67' & r_5 = -0.33' & r_8 = 2'\\
r_3 = 0.33' & r_6 = 0.67' & r_9 = -1'
\end{array}$$

From (6.6), we calculate an estimate of the variance of the equations:

$$S^2 = \frac{\sum_{i=1}^{n} r_i^2}{n - m + k} = \frac{\sum_{i=1}^{9} r_i^2}{9 - 2} = \frac{14.34}{7} = 2.05.$$

Now $D_{11} = 6$, $D_{22} = 6$, and (6.7) give

$$S^2(\tilde{a}) = S^2(\tilde{b}) = \frac{6}{27} \times 2.05 = 0.456, \qquad S(\tilde{a}) = S(\tilde{b}) = 0.675.$$

The conditional equations are equally accurate and the estimates $\tilde{a}$, $\tilde{b}$, and $\tilde{c}$ are equal to one another. Therefore, we can write immediately $S(\tilde{c}) = 0.675$. Finally, we obtain $\tilde{A} = 89°57.33'$, $\tilde{B} = 45°5.67'$, $\tilde{C} = 44°57.00'$, and $S\left(\tilde{A}\right) = S\left(\tilde{B}\right) = S\left(\tilde{C}\right) = 0.68'$.

We now construct the confidence interval for each angle based on Student's distribution. The number of degrees of freedom in this case is equal to $9 - 2 = 7$, and for $\alpha = 0.95$, Student's coefficient $t_{0.95} = 2.36$. Therefore, $u_{0.95} = 2.36 \times 0.68' = 1.6'$. Thus, we obtain finally

$$A(0.95) = 89°57.3' \pm 1.6', \qquad B(0.95) = 45°5.7' \pm 1.6',$$
$$C(0.95) = 44°57.0' \pm 1.6'.$$

In the above, the notation $A(0.95)$ means the value of A with confidence probability 0.95, the same for B and C.

Example 6.2. We shall study the example, which was presented at the beginning of this chapter, of combined measurements of the capacitance of two capacitors. The results of the direct measurement for the individual capacitors and for the two capacitors connected in parallel and in series are as follows:

$$x_1 = 0.2071\,\mu\text{F}, \qquad x_2 = 0.2056\,\mu\text{F},$$
$$x_1 + x_2 = 0.4111\,\mu\text{F}, \qquad \frac{x_1 x_2}{x_1 + x_3} = 0.1035\,\mu\text{F}.$$

The last equation is nonlinear. We expand it in a Taylor series, for which we first find the partial derivatives

$$\frac{\partial f}{\partial C_1} = \frac{C_2(C_1 + C_2) - C_1 C_2}{(C_1 + C_2)^2} = \frac{C_2^2}{(C_1 + C_2)^2}$$

and analogously

$$\frac{\partial f}{\partial C_2} = \frac{C_1^2}{(C_1 + C_2)^2}.$$

As $C_1 \approx x_1$ and $C_2 \approx x_2$, we can write

$$C_1 = 0.2070 + e_1, \qquad C_2 = 0.2060 + e_2.$$

Note that the above expressions use 0.2070 and 0.2060 instead of original values of 0.2071 and 0.2056. This simplifies the number manipulations without sacrificing the accuracy: because the values are close, we simply allocate the small discrepancies to e_1 and e_2, respectively.

The expansion into Taylor series is done for the point with the coordinates $C_{1,0} = 0.2070$ and $C_{2,0} = 0.2060$. We obtain

$$\begin{aligned}
&\frac{C_{1,0}C_{2,0}}{C_{1,0} + C_{2,0}} = 0.10325\\
&\left(\frac{\partial f}{\partial C_1}\right)_{C_{1,0},C_{2,0}} = \frac{0.206^2}{(0.207 + 0.206)^2} = 0.249\\
&\left(\frac{\partial f}{\partial C_2}\right)_{C_{1,0},C_{2,0}} = \frac{0.207^2}{(0.207 + 0.206)^2} = 0.251.
\end{aligned}$$

Thus, the nonlinear equation is thus linearized into $0.10325 + 0.249e_1 + 0.251e_2 = 0.1035$, and, setting $x_1 = C_1$ and $x_2 = C_2$, the system of conditional equations becomes

$$\begin{aligned}
1 \times e_1 + 0 \times e_2 &= 0.0001,\\
0 \times e_1 + 1 \times e_2 &= -0.0004,\\
1 \times e_1 + 1 \times e_2 &= -0.0019,\\
0.249e_1 + 0.251e_2 &= 0.00025.
\end{aligned}$$

We now calculate the coefficients of the normal equations

$$\begin{aligned}
&[xx] = 1 + 1 + 0.249^2 = 2.062, \qquad [xy] = 1 + 0.249 \times 0.251 = 1.0625,\\
&[yy] = 1 + 1 + 0.251^2 = 2.063, \qquad [xl] = -0.0004 - 0.0019 + 0.249\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times 0.00025 = -0.001\,738,\\
&[yl] = -0.0004 - 0.0019 + 0.251 \times 0.00025 = -0.002\,237.
\end{aligned}$$

The normal equations will be

$$\begin{aligned}
2.062e_1 + 1.0625e_2 &= -0.001\,738,\\
1.0625e_1 + 2.063e_2 &= -0.002\,237.
\end{aligned}$$

We now find the unknowns e_1 and e_2. According to (6.5), we calculate

$$D = \begin{vmatrix} 2.062 & 1.0625 \\ 1.0625 & 2.063 \end{vmatrix} = 3.125,$$

$$D_x = \begin{vmatrix} -0.001738 & 1.0625 \\ -0.002\,237 & 2.063 \end{vmatrix} = -0.001\,22,$$

$$D_y = \begin{vmatrix} 2.062 & -0.001\,738 \\ 1.0625 & -0.002\,237 \end{vmatrix} = -0.002\,75.$$

From here we find

$$e_1 = \frac{D_x}{D} = -0.000\,39, \qquad e_2 = \frac{D_y}{D} = -0.000\,88.$$

Therefore,

$$\tilde{C}_1 = 0.2070 - 0.000\,39 = 0.206\,61\,\mu\text{F},$$
$$\tilde{C}_2 = 0.2060 = 0.000\,88 = 0.205\,12\,\mu\text{F}.$$

We find the residuals of the conditional equations by substituting the estimates obtained for the unknowns into the conditional equations:

$$r_1 = 0.00049, \quad r_3 = -0.00063,$$
$$r_2 = 0.00058, \quad r_4 = 0.00048.$$

Now we can use formula (6.6) to calculate an estimate of the variance of the conditional equations:

$$S^2 = \frac{\sum_{i=1}^{4} r_i^2}{4-2} = \frac{120 \times 10^{-8}}{2} = 6 \times 10^{-7}.$$

The algebraic complements of the determinant D will be $D_{11} = 2.063$ and $D_{22} = 2.062$. As $D_{11} \approx D_{22}$,

$$S^2(\tilde{C}_1) = S^2(\tilde{C}_2) = \frac{D_{11}}{D} S^2 = \frac{2.063}{3.125} \times 6 \times 10^{-7} = 4 \times 10^{-7},$$
$$S(\tilde{C}_1) = S(\tilde{C}_2) = 6.3 \times 10^{-4}\,\mu\text{F}.$$

6.6 General Remarks on Determination of the Parameters in Formulas from Empirical Data

The purpose of almost any investigation in natural science is to find regularities in the phenomena in the material world, and measurements provide objective data for achieving this goal.

It is desirable to represent the dependencies between physical quantities determined from measurements in an analytic form, i.e., in the form of formulas. The initial form of the formulas is usually established based on an informal analysis of the collection of data obtained. One important prerequisite of the analysis is the assumption that the dependence sought can be expressed by a smooth curve; physical laws usually correspond to smooth curves. Once the form of the formula is chosen, its parameters are then found fitting the corresponding curve into the empirical data, and this is most often done by the method of least squares.

This problem is of great importance, and many mathematical and applied studies are devoted to it. We shall discuss some aspects of the solution of this problem that are related to the application of the method of least squares. The application of this method is based on the assumption that the acceptable optimality criterion for the parameters sought is that the sum of squares of the deviations of the empirical data from the curve obtained be minimized. This assumption is often justified, but not always.

For example, sometimes the curve must be drawn so that it exactly passes through all prescribed points; this is a natural requirement if the coordinates of the points are known to be exact. The problem is then solved by the methods of the interpolation approximation, and it is known that the degree of the interpolation polynomial will be one less than the number of fixed points. Sometimes the maximum deviation of the experimental data from the curve, rather than the sum of the squares of the deviations, is minimized.

As we have pointed out, however, most often the sum of the squares of the indicated deviations is minimized using the least squares method. For this purpose, all measured values for the quantities (in physically justified combinations) are substituted successively into the chosen formula, resulting in a system of conditional equations. The conditional equations are then used to construct the normal equations; the solution of the latter gives the values sought for the parameters. Next, substituting the values obtained for the parameters into the conditional equations, the residuals of these equations can be found and the standard deviation of the conditional equations can be estimated from them (assuming the equations are of equal accuracy).

It is significant that in this case, the standard deviation of the conditional equations is determined not only by the measurement errors but also by the imperfect structure of the formula chosen to describe the dependence sought. For example, it is well known that the temperature dependence of the electric resistance of many metals is reminiscent of a parabola. In engineering, however, it is often found that some sections of this dependence can be approximated by a linear function. The inaccuracy of the chosen formula, naturally, is reflected in the standard deviation of the conditional equations. Even if all experimental data were free of any errors, the standard deviation would still be nonzero. Thus, in this case, the standard deviation characterizes not only the error of the conditional equations, but also that the empirical formula adopted does not correspond to the true relation between the quantities.

It follows from this discussion that the estimates of the variances of the parameters obtained by the above method become virtual in the sense that they characterize not only the random spread in the experimental data, as usual, but also the inaccuracy of the approximation, which is nonrandom.

6.7 Construction of Transfer Functions of Measuring Transducers

We now turn to one particularly important application of the least squares method, the construction of the transfer functions (sometimes also referred to as calibration curves) for measuring transducers and instruments. These curves are a common way in which the results of the calibration of these devices are presented. We shall discuss the problem of constructing linear transfer functions, which are most often encountered in practice.

In a linear transfer function, the relation between a quantity y at the output of a transducer and the quantity x at its input is expressed by the dependence

$$y = a + bx. \tag{6.8}$$

When calibrating the transducer, the values of $\{x_i\}$, $i = 1, \ldots, n$, in the range $[x_{\min}, x_{\max}]$ are applied to its input, and the corresponding output values $\{y_i\}$ are found. Using these data, we have to estimate the coefficients a and b.

Let us start with the least-squares method. Equation (6.8) gives a system of n conditional equations

$$bx_i + a - y_i = r_i.$$

Following the least-squares scheme presented above, we obtain the system of normal equations

$$b\sum_{i=1}^{n} x_i^2 + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \quad b\sum_{i=1}^{n} x_i + na = \sum_{i=1}^{n} y_i. \tag{6.9}$$

The principal determinant of the system (6.9) will be

$$D = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2.$$

The determinant D_x is given by

$$D_x = \begin{vmatrix} \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix} = n\sum_{i=1}^{n} (x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i.$$

From here we find an estimate of the coefficient b:

$$\tilde{b} = \frac{D_x}{D} = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n} x_i^2 - n(\bar{x})^2}.$$

It is not difficult to show that

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{6.10}$$

and that

$$\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{6.11}$$

Then the expression for $\tilde{b}$ assumes the simpler form

$$\tilde{b} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}. \tag{6.12}$$

The determinant D_y is given by

$$D_y = \begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} y_i \end{vmatrix} = n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i .$$

Therefore,

$$\tilde{a} = \frac{D_y}{D} = \frac{n\bar{y}\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}\sum\limits_{i=1}^{n} x_i y_i}{n\sum\limits_{i=1}^{n} x_i^2 - n^2(\bar{x})^2}$$

Using the identity (6.11), we put the estimate $\tilde{a}$ into the form

$$\tilde{a} = \frac{\bar{y}\sum\limits_{i=1}^{n} x_i^2 - \bar{x}\sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \tag{6.13}$$

Relations (6.12) and (6.13) solve the problem of determining the transformation function

$$y = \tilde{a} + \tilde{b}x. \tag{6.14}$$

We now evaluate the uncertainty of the above solution. From the experimental data and the obtained estimates $\tilde{a}$ and $\tilde{b}$, we find the residuals of the conditional equations

$$r_i = \tilde{a} + \tilde{b}x_i - y_i.$$

Next, according to the general scheme of the least-squares method, we calculate the estimate of variance of the conditional equations using (6.6),

$$S^2 = \frac{\sum_{i=1}^{n} r_i^2}{n-2},$$

and estimates of the variances of $\tilde{a}$ and $\tilde{b}$ using (6.7). Finally, we find the confidence limits u_a and u_b, which represent the uncertainty of the two parameters. As pointed out above, the confidence limits are constructed based on Student's distribution with $n-2$ degrees of freedom in our case, because the confidence limits of two parameters are being determined.

The above confidence limits allow one to construct the so-called uncertainty band for the transfer function of the transducer. This band is depicted in Fig. 6.1. The band of uncertainty determines the range of possible transfer functions for the transducer.

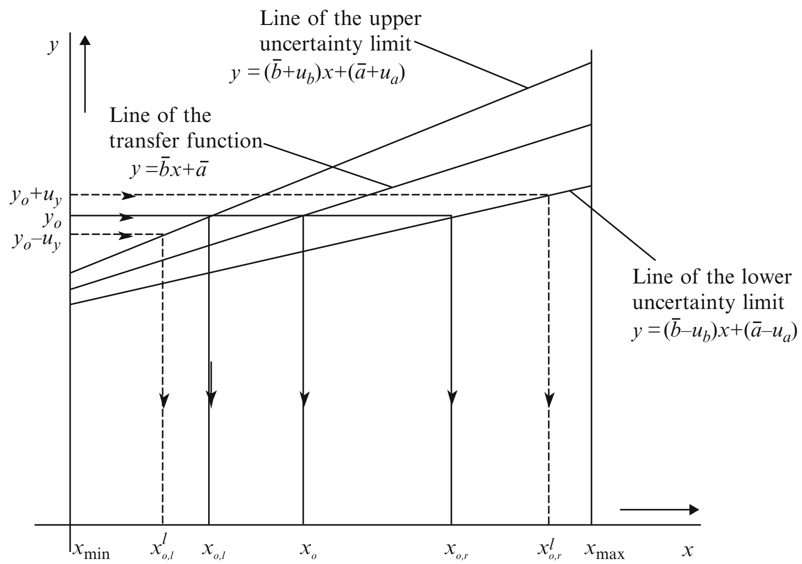


Fig. 6.1 Linear transfer function for the range $[x_{min}, x_{max}]$ and its band of uncertainty

It can be used to determine the accuracy of measurements obtained with the measuring transducer as follows.

When working with measuring transducers the dependence $x = f(y)$ and not $y = \varphi(x)$ is typically required: we need to obtain the value of the input signal by the observed value of the output signal. Consider a transducer with the band of uncertainty in Fig. 6.1 and let the observed signal be y_o. Assuming that the observed output value could be read precisely, the confidence interval for the input signal, $\left[x_{o,l}, x_{o,r}\right]$, is determined by the intersections of the horizontal line $y = y_o$ with the boundaries of the band of uncertainty.

If the output value itself is read with an uncertainty, $y_o \pm u_y$, then the confidence interval can be conservatively obtained as $\left[x'_{o,l}, x'_{o,r}\right]$ in Fig. 6.1. This confidence interval is conservative because is it not likely that both the output signal and the transfer function reach their respective boundary values simultaneously.

Note that the confidence intervals for the input value obtained above are not symmetrical around the "middle" value x_o given by the line of the transfer function. In practice, however, the band of uncertainty is narrow, and for narrow bands this asymmetry is negligible.

The least-squares method is not the only technique to construct a linear dependency between two measured quantities. In many cases, one can also build a linear dependency and its uncertainty band using the theory of indirect measurements. We discuss this last approach below.

During the calibration of transducers, it is common to obtain the output signal for the zero value of the input signal; this often corresponds to marking the initial value of the output indication of the transducer when no input signal is applied. Furthermore, this measurement can usually be viewed as precise compared to the other measurements: while other values of the input signal must be obtained from some device with certain accuracy, the absence of the signal corresponds to the true zero value. Then, for $x = 0$, (6.8) gives $\tilde{a} = y_0$, where y_0 is the corresponding output value.

Consider that we now have an estimate $\tilde{a}$ of the coefficient a. Then (6.8) can be transformed into the form

$$b = \frac{y - \tilde{a}}{x}.$$

This equation can be viewed as the measurement equation for the indirect measurement of the measurand b using the measuring arguments x and y. Because the values of y depend on the values of x, it is a dependent indirect measurement.

Calibration provides us with n pairs of x_i, y_i. Using the method of reduction, we transform this set of $\{x_i,\ y_i\}$ into a set $\{b_i\}$, $i = 1, \ldots, n$, which allows us to obtain the estimate of the coefficient b, $\tilde{b} = \bar{b}$, and its variance $S\left(\bar{b}\right)$. The uncertainty of coefficient b is determined using Student's distribution:

$$u_b = t_q S\left(\bar{b}\right),$$

where t_q is the Student coefficient for a given confidence probability and the degree of freedom $n - 1$. With this uncertainty, one can draw the transfer function and its band of uncertainty similar to Fig. 6.1. The only difference in this case is that the curves are constructed for interval $[0, x_{\max}]$ and all three curves converge to the same point $y = \tilde{a}$ on the y-axis.

We should note that the above application of the method of reduction assumes that all conditional equations are of equal accuracy, that is, all values of the input signal, $\{x_i\}$, are set with the same relative accuracy, and all values of the output signal, $\{y_i\}$, are measured also with the same relative accuracy. Otherwise calculations of the estimate $\tilde{b}$ and its variance would be more complex and less accurate (one would have to calculate $\tilde{b}$ as a weighted average of $\{b_i\}$; we omit further details).

Finally, it is useful to mention that during calibration, one should utilize diverse values of the input signal rather than perform repeated measurements of the output signal at the same value of the input. Indeed, in the latter case, the observed spread of values $\{b_i\}$ would characterize only one point in the transfer function and would not reflect the properties of the device in its entire range.

Chapter 7
Combining the Results of Measurements

7.1 Introductory Remarks

Measurements of the same quantity are often performed in different laboratories and, therefore, under different conditions and by different methods. Sometimes there arises the problem of combining these measurement data to find the most accurate estimate of the measured quantity.

In many cases, in the investigation of new phenomena, measurements of the quantities involved take a great deal of time. By grouping measurements performed over a limited time, intermediate estimates of the measurand can be obtained in the course of the measurements. It is natural to find the final result of a measurement by combining the intermediate results.

These examples show that the problem of combining the results of measurements is of great significance for metrology. At the same time, it is important to distinguish situations in which one is justified in combining results from those in which one is not justified in doing so. It is pointless to combine results of measurements of quantities that in their essence have different magnitude.

We should note that when comparing results of measurements, the data analysis is often performed based on the intuition of the experimenters without using formalized procedures. It is interesting that in the process, as a rule, the correct conclusions are drawn. On the one hand, this indicates that modern measuring instruments are of high quality and on the other hand that the experimenters, who by estimating the errors determine all sources of error, are usually highly qualified.

7.2 Theoretical Principles

The following problem has a mathematically rigorous solution. Consider L groups of measurements of the same quantity A. Estimates of the measurand $\bar{x}_1, \dots, \bar{x}_L$ were made from the measurements of each group, and

$$E[\bar{x}_1] = \cdots = E[\bar{x}_L] = A.$$

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*,
DOI 10.1007/978-1-4419-1456-9_7,

The variances of the measurements in each group $\sigma_1^2, \ldots, \sigma_L^2$ and the number of measurements in each group $n_1, \ldots, n_L$ are known. The problem is to find an estimate of the measured quantity based on data from all groups of measurements. This estimate is denoted as $\bar{\bar{x}}$ and is called the *combined average*. Because the combined average is commonly obtained as a linear combination of group averages, it is often referred to as the *weighted mean*.

We shall seek $\bar{\bar{x}}$ as a linear combination of $\{\bar{x}_j\}$, that is, as their weighted mean:

$$\bar{\bar{x}} = \sum_{j=1}^{L} g_j x_j. \tag{7.1}$$

Therefore, the problem reduces to finding the weights g_j. As $E\left[\bar{x}_j\right] = A$ for all j, and we obviously want $E\left[\bar{\bar{x}}\right] = A$, we obtain from (7.1)

$$E[\bar{\bar{x}}] = E\left[\sum_{j=1}^{L} g_j \bar{x}_j\right] = \sum_{j=1}^{L} g_j E[\bar{x}_j], \quad A = A\sum_{j=1}^{L} g_j.$$

Therefore,

$$\sum_{j=1}^{L} g_j = 1 \tag{7.2}$$

Next, we require that $\bar{\bar{x}}$ be an efficient estimate of A; that is, $V[\bar{\bar{x}}]$ must be minimum. $V[\bar{\bar{x}}]$ can be found using the formula

$$V[\bar{\bar{x}}] = V\left[\sum_{j=1}^{L} g_j \bar{x}_j\right] = \sum_{j=1}^{L} g_j^2 V[\bar{x}_j] = g_1^2\sigma^2(\bar{x}_1) + g_2^2\sigma^2(\bar{x}_2) + \cdots + g_L^2\sigma^2(\bar{x}_L). \tag{7.3}$$

We shall now find the weights g_j under which $V[\bar{\bar{x}}]$ reaches a minimum. Using the condition (7.2), we substitute $g_L = 1 - g_1 - g_2 - \cdots - g_{L-1}$ into (7.3), and then differentiate the resulting expression with respect to each g_j and equate each derivative to 0:

$$\begin{aligned}
2g_1\sigma^2(\bar{x}_1) - 2(1 - g_1 - g_2 - \cdots - g_{L-1})\sigma^2(\bar{x}_L) &= 0,\\
2g_2\sigma^2(\bar{x}_2) - 2(1 - g_1 - g_2 - \cdots - g_{L-1})\sigma^2(\bar{x}_L) &= 0,\\
&\cdots\\
2g_{L-1}\sigma^2(\bar{x}_{L-1}) - 2(1 - g_1 - g_2 - \cdots - g_{L-1})\sigma^2(\bar{x}_L) &= 0,
\end{aligned}$$

As the second term is identical in each equation, we obtain

$$g_1\sigma^2(\bar{x}_1) = g_2\sigma^2(\bar{x}_2) = \cdots = g_{L-1}\sigma^2(\bar{x}_{L-1}).$$

Furthermore, if instead of g_L we eliminated another weighting coefficient from (7.3), we would have included the similar term with g_L into the above relation.

Thus, we arrive at the following condition:

$$g_1\sigma^2(\bar{x}_1) = g_2\sigma^2(\bar{x}_2) = \cdots = g_L\sigma^2(\bar{x}_L),$$

or equivalently,

$$g_1 : g_2 : \cdots : g_L = \frac{1}{\sigma^2(\bar{x}_1)} : \frac{1}{\sigma^2(\bar{x}_2)} : \cdots : \frac{1}{\sigma^2(\bar{x}_L)}. \tag{7.4}$$

The relations (7.2) and (7.4) represent two conditions for the weights to compute the combined average. To find weight g_j, it is necessary to know either the variances of the arithmetic means or the ratio of the variances. If we have the variances $\sigma^2(\bar{x}_1)$, then we can set $g'_j = {}^1\!/_{\sigma^2(\bar{x}_1)}$. We then obtain

$$g_j = \frac{g'_j}{\sum\limits_{j=1}^{L} g'_j}. \tag{7.5}$$

As the weights are nonrandom quantities, it is not difficult to determine the variance for $\bar{\bar{x}}$. According to relation (7.3), we have

$$V[\bar{\bar{x}}] = \sum_{j=1}^{L} g_j^2 V[\bar{x}_j] = \frac{\sum\limits_{j=1}^{L}\left(g'_j\right)^2 V[\bar{x}_j]}{\left(\sum\limits_{j=1}^{L} g'_j\right)^2} = \frac{\sum\limits_{j=1}^{L}\left(\frac{1}{\sigma^2(\bar{x}_j)}\right)^2 \sigma^2(\bar{x}_j)}{\left(\sum\limits_{j=1}^{L}\frac{1}{\sigma^2(\bar{x}_j)}\right)^2} = \frac{1}{\sum\limits_{j=1}^{L}\frac{1}{\sigma^2(\bar{x}_j)}}. \tag{7.6}$$

Let us now consider an important particular case when the variances (7.6) of the measurements are the same for all groups, although their *estimates* might still be different because the number of observations in the groups may be different. In this case, one can combine the measurements of all groups into one large group of measurements. The number of measurements in the combined group is $N = \sum_{j=1}^{L} n_j$, and the combined average will be

$$\bar{\bar{x}} = \frac{\sum\limits_{j=1}^{L}\sum\limits_{i=1}^{n_j} x_{ji}}{N}. \tag{7.7}$$

Expanding the numerator gives

$$\begin{aligned}\bar{\bar{x}} &= \frac{(x_{11} + x_{12} + \cdots + x_{1n_1}) + (x_{21} + x_{22} + \cdots + x_{2n_2}) + \cdots}{N} \\ &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_L\bar{x}_L}{N} = \sum_{j=1}^{L} g_j\bar{x}_j,\end{aligned}$$

where g_j is the weight of the jth arithmetic mean:

$$g_j = n_j / N. \tag{7.8}$$

The standard deviation of the weighted mean in this case (i.e., when measurement results in each group have equal variances) can be estimated by considering the weighted mean as the average of the combined group of all the measurements:

$$S^2(\bar{\bar{x}}) = \frac{\sum\limits_{k=1}^{N} (x_k - \bar{\bar{x}})^2}{N(N-1)}.$$

We gather the terms in the numerator by groups

$$S^2(\bar{\bar{x}}) = \frac{\sum\limits_{j=1}^{L} \sum\limits_{i=1}^{n_j} (x_{ji} - \bar{\bar{x}})^2}{N(N-1)}$$

and perform simple transformations of the numerator to simplify the calculations:

$$\begin{aligned}
\sum_{j=1}^{L}\sum_{i=1}^{n_j}(x_{ji} - \bar{\bar{x}})^2 &= \sum_{j=1}^{L}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j + \bar{x}_j - \bar{\bar{x}})^2 \\
&= \sum_{j=1}^{L}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j)^2 + 2\sum_{j=1}^{L}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j)(\bar{x} - \bar{\bar{x}}) \\
&\quad + \sum_{j=1}^{L}\sum_{i=1}^{n_j}(\bar{x}_j - \bar{\bar{x}})^2.
\end{aligned}$$

The second term in the last expression is equal to zero, because by virtue of the properties of the arithmetic mean, $\sum_{i=1}^{n_j} \left(x_{ji} - \bar{x}_j\right) = 0$. For this reason,

$$S^2(\bar{\bar{x}}) = \frac{1}{N(N-1)} \left(\sum_{j=1}^{L}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^{L}\sum_{i=1}^{n_j}(\bar{x}_j - \bar{\bar{x}})^2 \right).$$

Note that

$$\sum_{i=1}^{n_j} \left(x_{ji} - \bar{x}_j\right)^2 = n_j \left(n_j - 1\right) S^2\left(\bar{x}_j\right),$$

where $S^2\left(\bar{x}_j\right)$ is the estimate of the variance of arithmetic mean of the jth group, or, equivalently,

$$S^2\left(\bar{x}_j\right) = \frac{1}{n_j\left(n_j - 1\right)} \sum_{i=1}^{n_j} \left(x_{ji} - \bar{x}_j\right)^2.$$

Further,

$$\sum_{i=1}^{n_j} \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2.$$

Thus, we obtain

$$S^2\left(\bar{\bar{x}}\right) = \frac{1}{N(N-1)}\left[\sum_{j=1}^{L}\left(n_j - 1\right) n_j S^2\left(\bar{x}_j\right) + \sum_{j=1}^{L} n_j \left(x_j - \bar{\bar{x}}\right)^2\right]. \tag{7.9}$$

Equation (7.9) can be expressed differently. Moving N in the denominator inside the square brackets, we have

$$S^2\left(\bar{\bar{x}}\right) = \frac{1}{N-1}\left[\sum_{j=1}^{L}\left(n_j - 1\right)\frac{n_j}{N} S^2\left(\bar{x}_j\right) + \sum_{j=1}^{L} \frac{n_j}{N} \left(\bar{x}_j - \bar{\bar{x}}\right)^2\right].$$

Finally, using (7.8), we obtain:

$$S^2\left(\bar{\bar{x}}\right) = \frac{1}{N-1}\left[\sum_{j=1}^{L} g_j \left(n_j - 1\right) S^2\left(\bar{x}_j\right) + \sum_{j=1}^{L} g_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2\right]. \tag{7.10}$$

The first term in the above formula characterizes the spread in the measurements within groups, and the second term characterizes the spread of the arithmetic means of the groups.

7.3 Effect of the Error of the Weights on the Error of the Weighted Mean

Looking at (7.1) determining the weighted mean, one would think that, because the weights g_j and the weighted values of $\bar{x}_j$ appear in it symmetrically, they must be found with the same accuracy. In practice, however, the weights are usually expressed by numbers with one or two significant figures. How is the uncertainty of the weights reflected in the error of the weighted mean?

We shall consider weights g_j in (7.1) to be fixed, constant values. In addition, as usual, we shall assume that the weights add up to one [that is, condition (7.2) holds]. This condition is also satisfied for the inaccurately determined weight estimates, that is, for $\tilde{g}_j$. Therefore,

$$\sum_{j=1}^{L} \Delta g_j = 0,$$

where Δg_j is the error in determining the weight g_j.

Assuming that the exact value of the weighted mean is y, we estimate the error of its estimate:

$$\Delta y = \sum_{j=1}^{L} \tilde{g}_j \bar{x}_j - \sum_{j=1}^{L} g_j \bar{x}_j = \sum_{j=1}^{L} \Delta g_j \bar{x}_j.$$

We shall express Δg_1 with the other errors:

$$\Delta g_1 = -(\Delta g_2 + \cdots + \Delta g_L)$$

and substitute it into the expression for y:

$$\Delta y = (\bar{x}_2 - \bar{x}_1)\Delta g_2 + (\bar{x}_3 - \bar{x}_1)\Delta g_3 + \cdots + (\bar{x}_L - \bar{x}_1)\Delta g_L$$

or in the form of relative error

$$\frac{\Delta y}{y} = \frac{g_2(\bar{x}_2 - \bar{x}_1)\frac{\Delta g_2}{g_2} + \cdots + g_L(\bar{x}_L - \bar{x}_1)\frac{\Delta g_L}{g_L}}{\sum\limits_{j=1}^{L} g_j \bar{x}_j}.$$

The errors of the weights $\Delta g_j / g_j$ are unknown. But let us assume that we can estimate their limits, and let $\Delta g / g$ be the largest absolute value of these limits. Replacing all relative errors $\Delta g_j / g_j$ with $\Delta g / g$, we obtain the upper limit of the relative error of the weighted mean:

$$\frac{\Delta y}{y} \le \frac{\Delta g}{g} \left(\frac{[|g_2(\bar{x}_2 - \bar{x}_1) + g_3(\bar{x}_3 - \bar{x}_1) + \cdots + g_L(\bar{x}_L - \bar{x}_1)|]}{\sum\limits_{j=1}^{L} g_j \bar{x}_j} \right).$$

The numerator on the right-hand side of the inequality can be put into the following form:

$$\begin{aligned} &g_2(\bar{x}_2 - \bar{x}_1) + g_3(\bar{x}_3 - \bar{x}_1) + \cdots + g_L(\bar{x}_L - \bar{x}_1) \\ &\quad = g_2\bar{x}_2 + g_3\bar{x}_3 + \cdots + g_L\bar{x}_L - (g_2 + g_3 + \cdots + g_L)\bar{x}_1. \end{aligned}$$

But $g_2 + g_3 + \cdots + g_2 = 1 - g_1$, so that

$$g_2(\bar{x}_2 - \bar{x}_1) + g_3(\bar{x}_3 - \bar{x}_1) + \cdots + g_L(\bar{x}_L - \bar{x}_1) = \sum_{j=1}^{L} g_j \bar{x}_j - \bar{x}_1 = y - \bar{x}_1.$$

Thus,

$$\frac{\Delta y}{y} \le \frac{\Delta g}{g} \frac{|y - \bar{x}_1|}{y}.$$

It is obvious that if the entire derivation is repeated, but in so doing the error not in the coefficient g_1 but in some other weight is eliminated, then a weighted value other than $\bar{x}_1$ will appear on the right-hand side of the inequality. Therefore, the above inequality holds for every $\bar{x}_j$; the obtained result can be represented in the form

$$\frac{\Delta\bar{\bar{x}}}{\bar{\bar{x}}} \le \frac{\Delta g}{g}\frac{|\bar{\bar{x}} - \bar{x}_j|}{\bar{\bar{x}}}.$$

This inequality shows that the error introduced into the weighted mean as a result of the error of the weights is many times smaller than the error of the weights itself. The cofactor $\left|\bar{\bar{x}} - \bar{x}_j\right|/\bar{\bar{x}}$ can be assumed to be of the same order of magnitude as the relative error of the measurement results $\bar{x}_j$ produced by each group. Thus, if this error is of the order of 0.01, then the error introduced into the weighted mean as a result of the error of the weights will be at least 100 times smaller than the latter.

7.4 Combining the Results of Measurements with Predominately Random Errors

We shall now study a scenario of combining measurement results where measurements in each group have negligibly small systematic errors. Each result being combined in this case is usually the arithmetic mean of the measurements in the corresponding group, and the differences between them are explained by the random spread of the averages of the groups.

Before attempting to combine these results, one must verify that the same quantity is measured in each case and there are no systematic shifts between the measurement results produced by each group. This verification is equivalent to checking that the true value of the measured quantity is the same for all groups and is accomplished by the methods presented in Chap. 3.

It is important to note that this verification can fail for two reasons: different quantities could have been measured in different groups or there are systematic shifts between the means of the groups. In the former case, it is pointless to combine the measurements. In the latter case the measurements can still be combined but with the help of another method, which we will discuss in the next section. The distinction between these two causes of verification failure must be clear from the physical essence of the measurement and its purpose; one cannot draw this distinction from statistical methods.

Only if the data pass the above verification can we combine the measurements by applying the approach from Sect. 7.2. Indeed, the absence or negligible size of the systematic errors is a necessary condition for the validity of this approach. One may notice that our verification only checks for the absence of the systematic shift between the groups, not the absence of the systematic errors themselves. This is inevitable; if measurements in all the groups have the same systematic error, this error is impossible to detect with statistical methods and it will also be present in the combined measurement result. Fortunately, this situation rarely occurs in practice. Recall that different groups of measurements are typically collected in different laboratories. Any systematic error that is so pervasive that it is the same across all the laboratories is likely to have been eliminated during calibration of the instruments involved.

The theory of calculating the weighted mean of several groups of measurements that we considered in Sect. 7.2 assumes that the variance of the measurement results in each group is known. However, the experimental data only allow one to obtain the estimates of these variances. Thus, one has to use the estimates in places of true variances throughout the calculations. In particular, the variance estimate of the weighted mean is computed by the following formula, modified from (7.6):

$$S^2\left(\bar{\bar{x}}\right) = \frac{1}{\sum\limits_{j=1}^{L} \frac{1}{S^2(\bar{x}_j)}}. \tag{7.11}$$

In the case of equal variances in all the groups, (7.9) and (7.10) already contain estimates of the group variance, and so these formulas can be used directly. Note that one can check if the estimates of the variances of measurement groups are the estimates of the same variance using the methods from Chap. 3.

Given this variance estimate, the uncertainty of the weighted mean can be calculated by considering the combination of the group averages as a linear indirect measurement and thus by applying (5.19) to calculate the effective degrees of freedom.

Example 7.1. The mass of some body is being measured. In one experiment, the value $\tilde{m}_1 = 409.52\,\text{g}$ is obtained as the arithmetic mean of $n_1 = 15$ measurements. The variance of the group of measurements is estimated to be $S_1^2 = 0.1\,\text{g}^2$. In a different experiment, the value $\tilde{m}_2 = 409.44\,\text{g}$ was obtained with $n_2 = 10$ and $S_2^2 = 0.03\,\text{g}^2$. It is known that the systematic errors of the measurements are negligibly small, and the measurement results in each experiment can be assumed normally distributed. It is necessary to estimate the mass of the body and the variance of the result using data from both experiments.

We shall first determine whether the unification is justified, that is, whether an inadmissible difference exists between the estimates of the measured quantity in each group. Following the method described in Sect. 3.7,

$$\begin{aligned}
S^2(\bar{x}_1) &= \frac{S_1^2}{n_1} = \frac{0.1}{15} = 0.0067, \quad S^2(\bar{x}_2) = \frac{0.03}{10} = 0.003,\\
S^2(\bar{x}_1 - \bar{x}_2) &= S^2(\bar{x}_1) + S^2(\bar{x}_2) = 0.0097,\\
S(\bar{x}_1 - \bar{x}_2) &= 0.098,\\
\bar{x}_1 - \bar{x}_2 &= \tilde{m}_1 - \tilde{m}_2 = 0.08.
\end{aligned}$$

Assuming that the confidence probability $\alpha = 0.95$, Table A.1 gives $z_{\frac{1+\alpha}{2}} = 1.96$. Then, $z_{\frac{1+\alpha}{2}} S(\bar{x}_1 - \bar{x}_2) = 1.96 \times 0.098 = 0.19$. As $0.08 < 0.19$, the unification is possible.

To decide if we can use the simpler method based on (7.8)–(7.10), we shall check whether both groups of observations have the same variance. We do so using Fisher's test from Sect. 3.7. We compute:

$$F = S_1^2/S_2^2 = 0.1 : 0.03 = 3.3.$$

The degrees of freedom are $\nu_1 = 14$ and $\nu_2 = 9$. We shall assume the significance level of 2%. Then, $q = 0.01$ and $F_q = 5$ (see Table A.5). As $F < F_q$, it can be assumed that the variances of the groups are equal.

We shall now find the weights of the arithmetic means. According to (7.8), we have $g_1 = 15/25 = 0.6$ and $g_2 = 10/25 = 0.4$. The weighted mean is $\bar{\bar{m}} = 0.6 \times 409.52 + 0.4 \times 409.44 = 409.49\,\text{g}$. Next we find $S\left(\bar{\bar{m}}\right)$. In accordance with (7.9), we have

$$\begin{aligned} S^2(\bar{\bar{m}}) &= \frac{1}{25 \times 24}(14 \times 0.1 + 9 \times 0.03 + 15 \times 0.03^2 + 10 \times 0.05^2) \\ &= 28 \times 10^{-4}\text{g}^2, \\ S(\bar{\bar{m}}) &= 5.3 \times 10^{-2}\text{g}. \end{aligned}$$

Having found the variance of the combined result, we can now calculate its uncertainty using Student's distribution with the effective degrees of freedom obtained from (5.19).

7.5 Combining the Results of Measurements Containing Both Systematic and Random Errors

In a general case, measurements within groups have not just random but also systematic error. The latter is typically a conditionally constant error or a sum of several conditionally constant errors. However, occasionally one may encounter absolutely constant systematic errors, such as methodological errors, as well. Let us start with considering measurements that do not have absolutely constant systematic errors.

Let us assume again that a quantity A is measured in L laboratories. Each laboratory produces the result $\bar{x}_j$ with error ζ_j $(j = 1, \ldots, L)$:

$$\bar{x}_j = A + \varsigma_j.$$

The error ζ_j is the sum of the conditionally constant error ϑ_j and random error ψ_j errors: $\zeta_j = \vartheta_j + \psi_j$. As discussed in Chap. 4 (Sect. 4.3), the conditionally constant error is modeled as a uniformly distributed random quantity with limits θ_j, which are estimated analytically from the specifications of the instruments and measurement conditions: $|\vartheta_j| \leq \theta_j$. We will assume that the mathematical expectation of this error is zero: $E[\vartheta_j] = 0$. We will also assume that θ_j is symmetrical about $\bar{x}_j$. Occasionally, one can encounter cases of asymmetrical limits; the methodology of handling this asymmetry is given in Chap. 4.

The random error ψ_j is assumed to be a centered quantity; that is, $E[\psi_j] = 0$. Thus, when there are no absolutely constant errors, we have $E\left[\bar{x}_j\right] = A$.

To allow the unification of measurement results, each laboratory must report the result itself, $\bar{x}_j$, along with the estimates of the variance of this result that is due to the random error, $S^2\left(\psi_j\right)$, and the limit of the conditionally constant systematic error θ_j. The former is calculated in the normal way:

$$S^2\left(\psi_j\right) = \frac{\sum\limits_{i=1}^{n_j}\left(x_{ji} - \bar{x}_j\right)^2}{n_j\left(n_j - 1\right)}.$$

The latter is equivalent to providing an estimate of the variance of this error, $S^2\left(\vartheta_j\right)$ since $S^2\left(\vartheta_j\right) = \theta_j^2/3$.

Similar to the case without systematic errors considered in Sect. 7.4, we will follow the theory of combining the results of measurements using the weighted mean while replacing variances with their estimates. As shown in Sect. 4.8, the estimate of the combined variance of the measurement result $\bar{x}_j$ is

$$S^2\left(\bar{x}_j\right) = S^2\left(\vartheta_j\right) + S^2\left(\psi_j\right). \tag{7.12}$$

Now, the weights of the results being combined can be derived from (7.2) and (7.4) by substituting the variances appearing in these relations with the estimates of these variances:

$$g_j = \frac{\dfrac{1}{S^2(\vartheta_j) + S^2(\psi_j)}}{\sum\limits_{j=1}^{L}\dfrac{1}{S^2(\vartheta_j) + S^2(\psi_j)}} \tag{7.13}$$

Knowing the weights, we can calculate the estimate of the combined result as the weighted mean of the results from each lab.

We shall now estimate the uncertainty of the weighted mean. In solving this problem, because the errors of the weights are insignificant (see Sect.7.3), we shall assume that the weights of the combined measurement results are exact. A necessary prerequisite to find the uncertainty is to estimate the standard deviation. In principle, we accomplish this by replacing variances in (7.5) with their estimates from (7.12). However, for subsequent calculations we will need the components of the combined standard deviation contributed by the random and conditionally constant systematic errors, denoted respectively as $S_\psi\left(\bar{\bar{x}}\right)$ and $S_\vartheta\left(\bar{\bar{x}}\right)$. Thus, we will compute these components and then obtain the overall standard deviation by combining these components rather than from (7.5) and (7.12).

Following the calculation procedure of Sect. 4.8, and taking into account the weights, $S_\psi\left(\bar{\bar{x}}\right)$ and $S_\vartheta\left(\bar{\bar{x}}\right)$ are computed as follows:

$$\begin{aligned} S_\psi\left(\bar{\bar{x}}\right) &= \sqrt{\sum_{j=1}^{L} g_j^2 S^2\left(\psi_j\right)} \\ S_\vartheta\left(\bar{\bar{x}}\right) &= \sqrt{\sum_{j=1}^{L} g_j^2 S^2\left(\vartheta_j\right)} \end{aligned}. \tag{7.14}$$

Now we can find the combined standard deviation of the weighted mean:

$$S\left(\bar{\bar{x}}\right) = \sqrt{S_\psi^2\left(\bar{\bar{x}}\right) + S_\vartheta^2\left(\bar{\bar{x}}\right)}. \tag{7.15}$$

To move from the combined standard deviation to the uncertainty of the weighted mean, according to (4.20), we must obtain coefficient t_c. This coefficient can be found from (4.22), which requires the coefficient t_ϑ for the systematic component of error and the quantile t_q of Student's distribution for the random component. To find t_ϑ we must first calculate the uncertainty of the systematic component. The easiest way to do it is by using (4.3) with weights:

$$u_\vartheta\left(\bar{\bar{x}}\right) = k\sqrt{\sum_{j=1}^{L} g_j^2 \theta_j^2}.$$

Coefficient k is determined by the desired confidence probability and is found from Table 4.1. Now we can find t_ϑ according to (4.21):

$$t_\vartheta = \frac{u_\vartheta\left(\bar{\bar{x}}\right)}{S_\vartheta\left(\bar{\bar{x}}\right)}.$$

Quantile t_q of Student's distribution can be found given the effective degrees of freedom using (5.19), which in this case obtains the form:

$$\nu_{\text{eff}} = \frac{\left[\sum_{j=1}^{L} g_j^2 S^2\left(\psi_j\right)\right]^2}{\sum_{j=1}^{L}\left(g_j^4 S^4\left(\psi_j\right)/\nu_j\right)},$$

where $\nu_j = n_j - 1$. Note that both t_ϑ and t_q must be obtained for the same confidence probability.

Now we can apply (4.22) to compute coefficient t_c

$$t_c = \frac{t_q S_\psi\left(\bar{\bar{x}}\right) + t_\vartheta S_\vartheta\left(\bar{\bar{x}}\right)}{S_\psi\left(\bar{\bar{x}}\right) + S_\vartheta\left(\bar{\bar{x}}\right)}$$

and, finally, obtain the uncertainty of the weighted mean:

$$u_c = t_c S\left(\bar{\bar{x}}\right).$$

We should say a few words on the possibility of absolutely constant systematic error. If among the groups being combined there is a group with such error, then the limit of this error must be re-calculated by taking into account the weight of this group. For instance if the only group with such error is group number 2 and its absolutely constant error is H_2 then the absolutely constant error of the weighted mean will be $H\left(\bar{\bar{x}}\right) = g_2 H_2$. If more than one group has such errors, their respective limits (again recalculated according to their groups' weights) are summed up arithmetically as in direct and indirect measurements. Then, the resulting limit is again summed up arithmetically with the confidence limit of the weighted mean computed using the methodology described here.

An example of a measurement where a weighted mean is used as the estimate of the measurand is a precise measurement of the activity of a source of alpha particles. A detailed treatment of this example is given in Chap. 8 (Sect. 8.8).

As a final note, when the results of measurements must be combined, it is always necessary to check the agreement between the starting data and the obtained result. If some contradiction is discovered, for example, the combined average falls outside the permissible limits of error of some group, then the reason for this must be determined and the contradiction must be eliminated. Sometimes this is difficult to do and may require special experiments. Great care must be exercised in combining the results of measurements because in this case information about the errors is employed to refine the result of the measurement and not to characterize its uncertainty, as is usually done.

7.6 Combining the Results of Single Measurements

Let us now consider an important special case when each group contains only a single measurement. In this case, the starting data include the estimates of the measurand and their inaccuracy. The inaccuracy can be given in the form of the limits of error or the uncertainty (confidence intervals) of the estimates. Our goal is to produce the weighted mean estimate of the measurand and its inaccuracy.

We begin with the case when the inaccuracies of individual measurements are given as limits of error. The error of each individual measurement is typically a conditionally constant systematic error, which, as discussed in Sect. 4.3, can be modeled as a random quantity with uniform distributions within its limits θ_j. Thus, its variance is related with the square of the limit of the distribution by a constant factor (the former is one-third of the latter). Therefore, the weights of these measurements can be computed to be reverse-proportionate to the squares of the corresponding limits of error θ_j rather than variances as in (7.4). Following the derivation of (7.4), we obtain:

$$g'_j = \frac{1}{\theta_j^2} \text{ and } \mathrm{g_j} = \frac{g'_j}{\sum_{j=1}^{L} g'_j}.$$

Having found the weights, we compute the weighted mean in the normal way. The inaccuracy of the weighted mean can be found using (4.3) while accounting for the weights of the terms, that is,

$$\theta_\alpha = k\sqrt{\sum_{j=1}^{L} g_j^2 \theta_j^2}.$$

We now turn to the case when the inaccuracy of individual measurements is represented in the form of uncertainties, or confidence intervals. Let $\theta_{j\alpha}$ be the uncertainty of j-th single measurement. We will assume that all the uncertainties were calculated for the same confidence probability α. Assume that uncertainty $\theta_{j\alpha}$ had been obtained from combining the m_j elementary errors involved in the j-th measurement using (4.3):

$$\theta_{j\alpha} = k_\alpha \sqrt{\sum_{i=1}^{m_j} \theta_{ji}^2} \quad \text{or} \quad \sum_{i=1}^{m_j} \theta_{ji}^2 = \frac{\theta_{j\alpha}^2}{k_\alpha^2}. \tag{7.16}$$

Formula (4.5) gives the expression for the variance of j-th measurement:

$$\sigma_j^2 = \frac{1}{3} \sum_{i=1}^{m_j} \theta_{ji}^2.$$

Replacing the sum with its expression given in (7.16), we obtain the estimate of the variance of j-th measurement:

$$S_j^2 = \frac{\theta_{j\alpha}^2}{3k_\alpha^2} = \frac{\sum\limits_{i=1}^{m_j} \theta_{ji}^2}{3}. \tag{7.17}$$

This formula indicates that all confidence limits are equally proportional to their corresponding variances. Then, the weights of the measurements can be computed to be reverse-proportionate to the squares of the corresponding confidence limits, analogously to the previous case when we used limits of error. And as in the previous case, we can now compute the weighted mean as the estimate of the measurand.

To calculate the inaccuracy of the weighted mean, note that its standard deviation can be computed from the standard deviations of its component as follows:

$$S\left(\bar{\bar{x}}\right) = \sqrt{\sum_{j=1}^{L} g_j^2 S_j^2},$$

or, utilizing (7.17),

$$S\left(\bar{\bar{x}}\right) = \frac{1}{\sqrt{3}}\sqrt{\sum_{j=1}^{L} g_j^2 \sum_{i=1}^{m_j} \theta_{ji}^2}. \tag{7.18}$$

To transition from the standard deviation to the confidence interval, note that the error of the weighted mean is a linear combination of all the elementary errors across all the single measurements. If the total number of the elementary errors, $\sum\limits_{j=1}^{L} m_j$, exceeds 4, which is practically always the case, we can consider the distribution of the weighted mean to be normal. Then, as we have seen multiple times already, the confidence limit of the overall result will be

$$u_\alpha = z_{\frac{1+\alpha}{2}} S\left(\bar{\bar{x}}\right).$$

In particular, $z_{\frac{1+\alpha}{2}} = 1.96$ for $\alpha = 0.95$ and $z_{\frac{1+\alpha}{2}} = 2.58$ for $\alpha = 0.99$.

We shall now discuss a particular case of single measurements when one quantity is measured independently with several instruments. We need to produce the combined measurement result and its inaccuracy.

Let the random errors of the instruments be small compared with the limit of permissible errors. First we consider the case when the permissible errors are the same and equal to Δ for all instruments. In this case, the problem can also be solved as follows. We choose the maximum and minimum indications of the instruments: $x_{\max}$ and $x_{\min}$. We verify that

$$(x_{\max} - x_{\min}) \leq 2\Delta.$$

If inequality (7.19) is not satisfied, then one of the instruments has an inadmissibly large error or the variation of some influence quantities is too large. The reason for this phenomenon must be determined and eliminated; that is, inequality (7.19) must be satisfied.

It is natural to take for the estimate of the measured quantity the center of the interval $x_{\max} - x_{\min}$:

$$\tilde{A} = \frac{x_{\max} + x_{\min}}{2}.$$

Figure 7.1 illustrates the indications $x_{\max}$ and $x_{\min}$ and shows the intervals corresponding to the limits of permissible errors $\pm\Delta$ of the corresponding instruments. The true value of the measured quantity must lie in the intersection of these two intervals; in the figure, this section is hatched. We will refer to this intersection as the *tolerance field*.

It follows from this figure that when the left boundary of the error interval of the upper device only abuts the right boundary of the error interval of the lower device, $x_{\max} = x_{\min} + 2\Delta$. This is one extreme case. The other extreme case is when $x_{\max} = x_{\min}$. It is easy to see that in both cases the error limits of the mean will be equal to $\pm\Delta$. Only when $x_{\max} = x_{\min} + \Delta$ will the limit error be $\pm\Delta/2$.

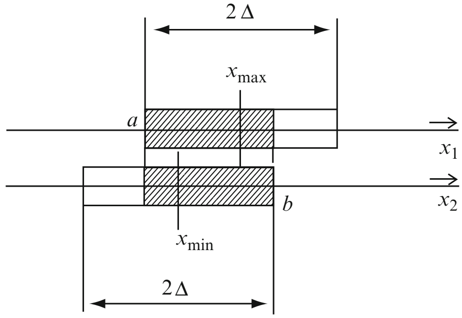


Fig. 7.1 The highest (x_{max}) and lowest (x_{min}) indications of the group of the instruments used to measure the same quantity; the interval of possible error of the combined measurement result is hatched

The likelihood of getting into this point is small. Furthermore, no matter how many instruments are used, the tolerance field is fully determined by the two instruments with the indications x_{max} and x_{min}. Thus, the parallel use of multiple equal accuracy instruments is not advisable.

Now we will show on a concrete example that there is no reason to measure the same quantity in parallel by several instruments of different accuracy. This will illustrate a well-known assumption of metrology that the accuracy of the measurement result is determined by the most accurate measuring instrument. Assume that the voltage of some source was measured simultaneously with three voltmeters having different accuracy but the same upper limit of the measurement range 15 V. The measurements were performed under reference conditions. Also, the voltage source has sufficient power for the consumption of the voltmeters to be considered negligible. The following results were obtained.

(1) Class 0.5 voltmeter: $U_1 = 10.05\,\text{V}$; the limit of permissible intrinsic error $\Delta_1 = 0.075\,\text{V}$.
(2) Class 1.0 voltmeter: $U_2 = 9.9\,\text{V}$; the limit of permissible intrinsic error $\Delta_2 = 0.15\,\text{V}$.
(3) Class 2.5 voltmeter: $U_3 = 9.7\,\text{V}$, the limit of permissible intrinsic error $\Delta_3 = 0.375\,\text{V}$.

As the measurements were performed under reference conditions, we shall assume that the limits of permissible intrinsic error of the instruments are equal to the limits of the errors of measurement.

Assume that the errors of the instruments of each type have a uniform distribution. Then

$$\sigma_i = \Delta_i/\sqrt{3}.$$

We will now combine these individual measurements into the overall result. We shall find the weights of the individual results based on the limits of intrinsic error of the instruments:

$$g_1' = \frac{1}{\Delta_1^2} = \frac{1}{0.25} = 4, \quad g_2' = \frac{1}{\Delta_2^2} = 1, \quad g_3' = \frac{1}{\Delta_3^2} = \frac{1}{6.25} = 0.16.$$

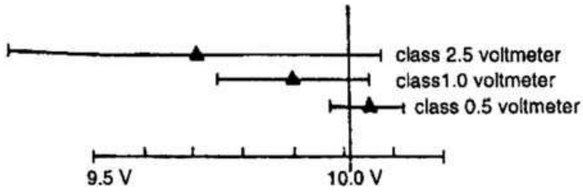


Fig. 7.2 The possible indications of voltmeters accuracy classes 2.5, 1.0 and 0.5 obtained in measurements of the same voltage and the intervals of their permissible errors; the weighted mean value is shown by the vertical line

From here,

$$g_1 = \frac{g'_1}{\sum_{i=1}^{3} g'_i} = \frac{4}{5.16} = 0.77,$$

$$g_2 = \frac{g'_2}{\sum_{i=1}^{3} g'_i} = \frac{0.20}{5.16} = 0.20, \quad g_3 = \frac{g'_3}{\sum_{i=1}^{3} g'_i} = \frac{0.16}{5.16} = 0.03.$$

Now we find the weighted mean

$$\tilde{U} = \sum_{i=1}^{3} g_i U_i = 0.77 \times 10.05 + 0.2 \times 9.9 + 0.03 \times 9.7 = 10.01\,\text{V}.$$

The confidence limits of the error in the weighted mean can be found from (4.3) with added weights:

$$\begin{aligned}\Delta\tilde{U} &= k\sqrt{\sum_{i=1}^{3} g_i^2 \Delta_i^2} \\ &= k\sqrt{0.77^2(7.5 \times 10^{-2})^2 + 0.2^2(15 \times 10^{-2})^2 + 0.03^2 \times 0.375^2} \\ &= k\sqrt{(33 + 9 + 1.3) \times 10^{-4}} = 0.066k.\end{aligned}$$

Assuming, as usual, $\alpha = 0.95$, we take $k = 1.1$ and find $\Delta\tilde{U} = 0.07\,\text{V}$.

Figure 7.2 plots the indications of all three instruments, with the limits of permissible error of the instruments marked. The vertical line indicates the value obtained for the weighted mean. This value remained in the error interval of the most accurate result, but it was shifted somewhat in the direction of indications of the less accurate instruments; this is natural. As we see the limits of error of the result decreased insignificantly compared with the error of the most accurate term.

Chapter 8
Examples of Measurements and Measurement Data Processing

8.1 Voltage Measurement with a Pointer-Type Voltmeter

Our first example concerns a measurement of voltage with a pointer-type voltmeter. Such a measurement clearly represents an example of a direct measurement. We shall study several examples of such measurements with a Class 1.0 pointer-type DC voltmeter that operates using the energy of the source of the voltage being measured. Note that the energy consumption by the voltmeter causes interaction between the voltmeter and the object under study.

Let the voltmeter have the following characteristics:

1. The upper limits of measurement ranges are 3 V, 7.5 V, 15 V, 30 V, and so on, up to 300 V.
2. The scale of the instrument has 75 graduations and starts at the 0 marker.
3. The limits of permissible intrinsic error are $\pm 1.0\%$ of a span (it is a fiducial error).
4. Full deflection of the pointer corresponds to the current of 15×10^{-6}A $\pm$ 1%.
5. Reference conditions include temperature of $+20 \pm 5$°C and the requirement that the measurement be performed with the instrument positioned horizontally.
6. Additional errors are as follows. A deviation of the temperature from the reference range causes the indications of the instrument to change by not more than $\pm 1.0\%$ for each 10°C change in temperature. Inclination of the instrument by 5° from the horizontal position changes the indications by not more than $\pm 1\%$ of the measurement range employed.

8.1.1 *A Priori Estimation of Measurement Inaccuracy*

Suppose that quality assurance of a piece of equipment involves measuring the voltage on certain pairs of points in its electrical schema. We can represent this unit as an equivalent source of voltage with EMF E and output resistance R connected serially to the voltmeter. The source resistance R in one case is equal to about 10 k Ω and in

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*, DOI 10.1007/978-1-4419-1456-9_8, © Springer Science+Business Media, LLC 2010

all other cases does not exceed $1\,\mathrm{k}\,\Omega$. The temperature of the medium can change from $+10°$C to $+25°$C. The slope relative to the horizontal position does not exceed $5°$. We are required to estimate the measurement uncertainty. The uncertainty must be expressed in the relative form.

Before the measurement, the value of the measured quantity is unknown. It will supposedly be less than 3 V. Considering the measurement ranges of the voltmeter, we note that there is an overlap of 0.4–0.5 between any two consecutive ranges. For example, the smallest range (3 V) represents 0.4 of the next higher range (since $3\,\mathrm{V}/7.5\,\mathrm{V} = 0.4$); the next range (7.5 V) represents 0.5 of the next range, and so on. Thus, whenever the voltmeter indication drops below 0.4–0.5 of a given range limit, one should switch to the preceding, lower, range. Following this logic, we shall assume that if the measured voltage is less than $0.4 \times 3\,\mathrm{V} = 1.2\,\mathrm{V}$, then a different voltmeter must be used.

Assume that the 3 V range is to be used (other ranges are treated similarly). In this range, the largest relative error will occur when a voltage at the low end of this range, or around 1.2 V, is being measured. The error will have to be estimated for this worst case.

The sources of error are as follows:

1. The intrinsic error of the voltmeter
2. The reading error
3. The temperature error
4. The error introduced by the inclination of the instrument
5. The error from the limited internal resistance of the voltmeter

The error from the limited resistance of the voltmeter is absolutely constant for each unit being tested. The other errors listed above are conditionally constant. We shall now estimate these errors.

1. Intrinsic error θ_{in}. Its limits will be

$$\theta_{\text{in}} = \pm 1\% \times \frac{1}{0.4} = \pm 2.5\%, \quad |\theta_{\text{in}}| = 2.5\%.$$

2. Reading error θ_r. This error does not exceed 0.25 of a graduation. When measuring 1.2 V at the limit 3 V, and with 75 graduations of the scale, this gives

$$\theta_r = \pm 0.25 \times \frac{3 \times 100\%}{75 \times 1.2} = \pm 0.83\%, \quad |\theta_r| = 0.83\%.$$

3. Additional temperature error θ_T. The maximum deviation of the temperature from the normal value is $(20 - 5) - 10 = 5°$C. Therefore,

$$\theta_T = \pm 1\% \times \frac{5}{10} = \pm 0.5\%, \qquad |\theta_T| = 0.5\%.$$

4. The additional error θ_l. Because of the 5° inclination of the instrument, the additional error when measuring 1.2 V will be

$$\theta_l = \pm 1\% \times \frac{3}{1.2} = \pm 2.5\%, \qquad |\theta_l| = 2.5\%.$$

5. The error H_R from the limited internal resistance of the voltmeter. The internal resistance of the voltmeter at the limit 3 V is

$$R_V = \frac{3}{15 \times 10^{-6}} = 2 \times 10^5 \Omega.$$

The indications of the voltmeter correspond to the voltage on its terminals. This voltage U is less than the EMF E in the circuit:

$$U = \frac{R_V}{R_V + R} E.$$

The error then is

$$H_R = \frac{U - E}{E} = \frac{-R_V}{R_V + R}.$$

The worst case occurs with the source resistance $R = 10\,\text{k}\,\Omega$, in which case this error becomes

$$H_R = \frac{-10 \times 10^3}{10 \times 10^3 + 2 \times 10^5} \times 100 = -4.8\%.$$

If the source resistance is $1\,\text{k}\,\Omega$, then $H_R = -0.5\%$.

Let us now add all conditionally constant errors. We shall use (4.3), and we shall assume that $\alpha = 0.95$:

$$u_{0.95} = 1.1\sqrt{2.5^2 + 0.83^2 + 0.5^2 + 0.25^2} = 4\%.$$

We now take into account the absolutely constant error. Its limits are

$$H_{Rl} = -4.8\%, \; H_{Rr} = -0.5\%,$$

but they are not known accurately enough to eliminate them by introducing the correction. Therefore, in accordance with (4.16), we obtain the overall limits of error:

$$\Delta_{r.0.95} = -0.5 + 4 = +3.5\%, \qquad \Delta_{l,0.95} = -4.8 - 4.0 = -8.8\%$$

Thus, the absolute value of error of the planned measurement will not exceed $\sim$10%.

8.1.2 Universal Estimation of Measurement Inaccuracy

We shall now estimate the measurement error in the example examined above, assuming that the measurement has already been made. The significant difference from the previous case is that now we have an estimate of the measured quantity.

Assume the case with source resistance $R = 10\,\mathrm{k}\Omega$ and let the indication of the voltmeter be 62.3 graduations. Hence, the voltage indicated by the voltmeter is

$$U = 62.3\frac{3}{75} = 2.492\,\mathrm{V}.$$

Suppose we found out that $R = 10\,\mathrm{k}\,\Omega \pm 0.5\%$. The error H_R was calculated above: $H_R = -4.8\%$. Now we can introduce the correction C_R:

$$C_R = +4.8 \times 10^{-2} \times 2.492 = +0.120\,\mathrm{V}.$$

Taking the correction into account, we obtain

$$U' = U + C_R = 2.612\,\mathrm{V}.$$

The error of the correction is determined by the errors of the values of the voltmeter resistance R_V and the source resistance R. We shall establish the relation between them.

$$C_R = -H_R U = \frac{R}{R + R_V}U = \frac{R}{R + R_V} \times \frac{R_V}{R + R_V}E = \frac{R/R_V}{(1 + R/R_V)^2}E.$$

To simplify the notation, let $x = R/R_V$. Then

$$C_R = \frac{x}{(1 + x)^2}E.$$

We now construct the differential relations:

$$dx = \frac{1}{R_V}dR - \frac{R}{R_V^2}dR_V = x\left(\frac{dR}{R} - \frac{dR_V}{R_V}\right),$$

$$dC_R = E\left(\frac{dx}{(1 + x)^2} - \frac{2x\,(1 + x)\,dx}{(1 + x)^4}\right) = E\frac{1 - x}{(1 + x)^3}dx,$$

$$dC_R = E\frac{x\,(1 - x)}{(1 + x)^3}\left(\frac{dR}{R} - \frac{dR_V}{R_V}\right).$$

In the relative form, transforming from differentials to increments, we obtain

$$\theta_C = \frac{\Delta C_R}{C_R} = \frac{1 - x}{1 + x}\left(\frac{\Delta R}{R} - \frac{\Delta R_V}{R_V}\right).$$

The above formula suggests that there are two components in the correction error due to R and R_V, respectively. We can express these components in a relative form as:

$$\theta_{C1} = \frac{1-x}{1+x}\theta_R, \quad \theta_{C2} = \frac{1-x}{1+x}\theta_{R_V},$$

where θ_R and θ_{R_V} are the relative errors of the outside resistance and voltmeter input resistance. As ΔR and ΔR_V are independent, we shall regard each component of error of the correction as an elementary error of measurement. Obviously, both components are conditionally constant.

Recall that the limits of the error of the source resistance R are known to be $\pm 0.5\%$. Therefore,

$$|\theta_{C1}| = \left(\frac{1-x}{1+x}\right) 0.5\% = 0.9 \times 0.5\% = 0.45\%.$$

The limits of error of the internal resistance of the voltmeter are determined by the voltmeter class. Since ours is a voltmeter of Class 1, these limits are equal to $\pm 1\%$. Therefore, because $x = 5 \times 10^{-2}$ for the values of R and R_V,

$$|\theta_{C2}| = \left(\frac{1-x}{1+x}\right) 1\% = 0.9 \times 1\% = 0.9\%.$$

The limits of the remaining errors are as follows:

$$|\theta_{\text{in}}| = 1\% \times 75/62 = 1.2\%$$
$$|\theta_r| = \frac{0.25 \times 100\%}{62} = 0.4\%$$
$$|\theta_T| = 0.5\%$$
$$|\theta_l| = 1\% \times 75/62 = 1.2\%.$$

These elementary errors can be assumed to be conditionally constant. According to (4.3), for $\alpha = 0.95$, we obtain

$$u_{0.95} = 1.1\sqrt{0.9^2 + 0.45^2 + 1.2^2 + 0.4^2 + 0.5^2 + 1.2^2} = 2.3\%.$$

When the result of the measurement is written in accordance with its uncertainty, only three significant figures can be retained:

$$\tilde{U} = 2.61\text{V}, \quad u = \pm 2.3\%(0.95).$$

Alternatively, the result can be represented as follows:

$$U_{0.95} = 2.61\text{V} \pm 2.3\%, \text{ or } U_{0.95} = (2.61\text{V} \pm 0.06)\text{ V}.$$

8.1.3 Individual Estimation of Measurement Inaccuracy

The largest elementary errors in the previous section were θ_{C2}, θ_{in}, and θ_l. How can they be reduced? The first two can be reduced by taking into account the individual properties of the voltmeter, if the voltmeter has a table of corrections from a recent calibration test. Assume that, for the 3 V measurement range, the correction is $+0.3$ graduations at marker 60, and $+0.2$ graduations at marker 70. It can then be assumed that the correction to the indication at 62.3 graduations is also equal to $+0.3$ graduations. Therefore,

$$C_{\text{in}} + 0.3 \times \frac{3}{75} = +0.012\,\text{V}.$$

Taking this correction into account, the voltmeter indication gives

$$U' = 2.492 + 0.012 = 2.504\,\text{V}.$$

We shall assume that the limits of error in determining the correction, i.e., the calibration errors, are known and are equal to $\pm 0.2\%$. Converting to the indication of the instrument, we obtain

$$|\theta_{\text{in}}| = 0.2 \times 75/62 = 0.24\%.$$

With this correction, we have eliminated the systematic component of the error of the voltmeter. The random component, however, remains, and it must be taken into account. The dead band in indicating electric measurement instruments can reach a value coinciding with the class designation of the instrument. In our case, this value is 1% of 3 V. The random error does not exceed half the dead band. Thus, the limits of random error are equal to

$$|\Psi| = 0.5 \times 1\% \times \frac{75}{62} = 0.6\%$$

The distribution of the random error in our case, once its limits have been estimated, can be assumed to be uniform, as also the distributions of other conditionally constant elementary errors.

The input resistance of the voltmeter can be measured. Assume that this measurement has been done, and $R_V = 201.7\,\text{k}\,\Omega \pm 0.2\%$. Then

$$H_R = \frac{-10 \times 10^3 \times 100}{(10 + 201.7) \times 10^3} = -4.72\%.$$

The correction will then be

$$C_R = +4.72 \times 10^{-2} \times 2.504 = +0.118\,\text{V}.$$

Taking the correction C_R into account, we obtain

$$U'' = 2.504 + 0.118 = 2.622\,\text{V}.$$

The limits of the elementary error θ_{C1} do not change, but θ_{C2} will now become smaller due to not knowing the exact input resistance of the voltmeter:

$$|\theta_{C1}| = 0.45\%, \quad |\theta_{C2}| = 0.9 \times 0.2\% = 0.18\%.$$

The error θ_l can be reduced by taking greater care in positioning the instrument horizontally. Assume that the deviation from the horizontal position does not exceed $\pm 2°$. Then

$$|\theta_l| = 1 \times 2/5 \times 75/62 = 0.48\%.$$

The temperature error and the reading error will remain the same.

Let us calculate the uncertainty again for $\alpha = 0.95$:

$$u_{0.95} = 1.1\sqrt{0.24^2 + 0.6^2 + 0.18^2 + 0.45^2 + 0.48^2 + 0.5^2 + 0.4^2} = 1.2\%.$$

We now write the result of the measurement as follows:

$$\tilde{U} = 2.62\,\text{V}, \quad u = \pm 1.2\%(0.95),$$

or alternatively,

$$U_{0.95} = 2.62\,\text{V} \pm 1.2\%, \;\text{ or }\; U_{0.95} = (2.62\,\text{V} \pm 0.03)\ \text{V}.$$

This example illustrates clearly how the measurement uncertainty decreases as one moves from a priori to a posteriori estimation and then from universal to individual error estimation.

8.2 Voltage Measurement with a Potentiometer and a Voltage Divider

Potentiometers with manual control are highly accurate and universal. For these reasons, they are frequently used in scientific laboratories, although they have started to be displaced by digital multirange voltmeters in recent years. The latter are in essence automated potentiometers.

A voltage measurement with a potentiometer requires a two-phase measurement procedure. First, a standard cell is connected to the potentiometer, and the current through the potentiometer is adjusted using the potentiometer's set of accurate measuring resistors so that the voltage drop on the section of the circuit with these resistors would balance the EMF of the standard cell. Next, a special potentiometer switch is used to disconnect the standard cell, and we connect the voltage to be measured to the potentiometer circuit.

When the voltage to be measured exceeds the range of the potentiometer, a voltage divider can be used, which allows only a known fraction of the voltage to be applied to the potentiometer. We should point out that a voltage divider contains electrical resistors and thus consumes a certain amount of power from the voltage source to which it connects. For this reason, a voltage divider can only be used if the power it consumes is so low that the resulting affect on the measured voltage is negligible. We assume that this is the case in our example.

The measurement of voltage with a potentiometer is a direct measurement. However, when the errors of the potentiometer and the errors of the standard cell are rated separately, and when a voltage divider is involved, the error produced by such a chain of measuring instruments is estimated with methods that are specifically designed for indirect measurements. We discussed these methods in Chap. 5. Here, we shall consider an example of a single measurement with individual inaccuracy estimation.

To be specific, we will consider the measurement of voltage using a class 0.005 potentiometer, a class 0.005 voltage divider, and a standard cell with voltage accuracy of $\pm 10\,\mu\text{V}$. In particular, we will consider a P309 potentiometer and P35 voltage divider, which were manufactured in the former USSR. The measuring resisters in P309 potentiometer are organized in six blocks called decades. Each decade produces certain decimal digits in the measurement result. For example, if the measured voltage is 1.256316 V, the digits "1.2 V" are produced by indication "12" of decade "×100 mV," the digit "0.05 V" by indication "5" of decade "×10 mV," and so on.

Let the current through the potentiometer be I_p and the resistance of the section of the circuit with the accurate resistors after the adjustment in the first phase be R_{sc}. Since the voltage drop on the section of the circuit with the resistance R_{sc} balances the EMF of the standard cell, U_{sc}, we have in this case:

$$I_p = U_{\text{sc}}/R_{\text{sc}}.$$

When the standard cell is disconnected and a certain voltage, U_p, is connected to the potentiometer circuit, a fraction of the resistors of the potentiometer is introduced into the comparison circuit such that the voltage drop on their resistance R_p would compensate U_p; i.e., $U_p = I_p R_p$. Then

$$U_p = \frac{R_p}{R_{\text{sc}}} U_{\text{sc}},$$

and knowing the EMF of the standard cell and the ratio R_p/R_{sc}, we can find U_p. Finally, assuming that the division coefficient of the voltage divider is equal to K_d, the voltage to be measured, U, is determined from the formula $U = K_d U_p$. Therefore, we can write the measurement equation in this measurement in the form:

$$U = K_d \frac{R_p}{R_{sc}} U_{sc}. \tag{8.1}$$

The indications of the potentiometer are proportional to R_p, but its error is determined not by the errors of the resistances R_p and R_{sc}, but by the error of the ratio R_p/R_{sc}. The uncertainty associated with the operations of comparing the voltages can be neglected, because the smoothness of the resistance regulation in the potentiometer and the sensitivity of its zero indicator were designed specifically to keep this uncertainty extremely small compared to other errors.

The potentiometer has six decades and a built-in self-balancing amplifier. The limit of permissible error as a function of the measured voltage U_p is calculated using the formula (given in the manufacturer's documentation):

$$\Delta U_p = \pm \left(50U_p + 0.04\right) \times 10^{-6}\,\text{V}.$$

The error of the potentiometer does not exceed the above limits if the ambient air temperature ranges from $+15$ to $+30°$C and differs by not more than 2.5°C from the temperature at which the measuring resistors of the potentiometer were adjusted (the P309 potentiometer has built-in calibration and adjusting systems).

The EMF of the standard cells can be determined with an error of $\pm 10\,\mu$V that in relative form is $\pm 1 \times 10^{-3}\%$. The effect of the temperature is taken into account using a well-known formula, which describes accurately the temperature dependence of the EMF in a standard cell. Thus, temperature does not introduce additional errors to the EMF of the standard cell.

Assume that in three repeated measurements of certain voltage, performed using a voltage divider whose voltage division ratio was set to 1:10, the following potentiometer indications were obtained:

$$x_1 = 1.256316\,\text{V}, x_2 = 1.256321\,\text{V}, x_3 = 1.256318\,\text{V}.$$

The limit of permissible error of the potentiometer in this case is

$$\Delta U_p = \pm\,(50 \times 1.26 + 0.04) \times 10^{-6} = \pm 63\,\mu\,\text{V}.$$

For this reason, the difference of $5\,\mu$V between the results of the three observations above can be regarded as resulting from the random error of the measurement, and the magnitude of this error is negligible. In the calculation, therefore, any one of these results or their average value can be used.

Assume that in the process of adjusting the measuring resistors before the measurement, the corrections of the higher order decades were estimated. Let the correction for the indication "12" of the decade "$\times$100 mV" be $+15 \times 10^{-6}$ V, and the correction for the indication "5" of the decade "$\times$10 mV" be -3×10^{-6} V. Each correction is determined with an error of $\pm 5 \times 10^{-8}$ V.

The corrections for the other decades are so small that they are of no interest. Indeed, the indication of all the remaining decades is 0.0063 V; the limit of permissible error corresponding to this indication in accordance with the formula given above is

$$\Delta U_p = \pm(50 \times 0.0063 + 0.04) \times 10^{-6} = \pm 0.32 \times 10^{-6}\,\mathrm{V}.$$

This error is already two orders of magnitude smaller than the permissible error of the higher decades, and it can be neglected without further corrections.

Further, it is necessary to take into account the possible change in the air temperature in the room. If this change falls within permissible limits, then according to the specifications of the potentiometer, the error can change approximately by one-forth of the permissible limit, i.e., by 16 μV.

We shall take for the result the average value of the observations performed, correcting it by the amount $C = (15 - 3) \times 10^{-6} = 12 \times 10^{-6}\,\mu\mathrm{V}$:

$$U_p = \bar{x} = 1.256\,318 + 0.000\,012 = 1.256\,330\,\mathrm{V}.$$

The errors of the potentiometer, which enter into this result, include the error due to temperature ($\pm 16 \times 10^{-6}$ V), the error of correction of the higher decades ($\pm 5 \times 10^{-8}$ V), and the error due to the lower decades ($\pm 0.32 \times 10^{-6}$ V). Clearly, these errors are dominated by the error due to temperature, and the remaining errors can be neglected. Thus, the limits of error of the potentiometer are

$$\theta_p = \pm 16 \times 10^{-6}\,\mathrm{V}.$$

Next, we must estimate the errors from the standard cell and the voltage divider. The error of the class 0.005 voltage divider can reach $5 \times 10^{-3}\%$. But the actual division coefficient of the divider can be found and taken into account, which is precisely what we must do in the case at hand. In the given measurement, assume that this coefficient has been found to be $K_d = 10.0003$ and the error in determining K_d falls within the range $\pm 2 \times 10^{-3}\%$.

Finally, the discrepancy between the real and the nominal value of the EMF of the standard cell falls within the limits of error of the standard cell ($\pm 10\,\mu$V).

We estimate the voltage being measured U as

$$\tilde{U} = K_d U_p = 10.0003 \times 1.256330 = 12.56368\,\mathrm{V}.$$

To estimate the measurement error, we shall use the standard trick. First, we shall take the logarithm of the measurement (8.1). Then we find the differentials of both sides of the equation, and neglecting errors that are second-order infinitesimals, we replace the differentials by the increments. This process gives

$$\frac{\Delta U}{U} = \frac{\Delta K_d}{K_d} + \frac{\Delta\left(R_p/R_{\mathrm{sc}}\right)}{R_p/R_{\mathrm{sc}}} + \frac{\Delta U_{\mathrm{sc}}}{U_{\mathrm{sc}}}.$$

For the terms on the right side of the above formula, we only have estimates of the limits, and not the values of the errors. Thus, we shall estimate the limits of the measurement error on the left side. We can use formula (4.3) for this purpose. First, all components must be represented in the form of relative errors. The limits of the relative error of the potentiometer, in percent, will be

$$\theta_p = \pm \frac{16 \times 10^{-6} \times 100}{1.26} = \pm 1.3 \times 10^{-3}\%.$$

The limits of the relative error of the voltage divider were estimated directly as $\theta_K = \pm 2 \times 10^{-3}\%$. The limits of error in determining the EMF of the standard cell in the form of a relative error are known:

$$\theta_{\mathrm{sc}} = \pm 1 \times 10^{-3}\%.$$

We now find the limit of the measurement error according to (4.3):

$$\theta_\alpha = k\sqrt{1.3^2 + 2^2 + 1^2} \times 10^{-3} = k \times 2.6 \times 10^{-3}\%.$$

Let $\alpha = 0.95$. Then $k = 1.1$ and

$$\theta_{0.95} = 1.1 \times 2.6 \times 10^{-3} = 2.9 \times 10^{-3} \approx 3 \times 10^{-3}\%.$$

Finally, we must check the number of significant figures in the result of measurement. To this end, we shall express the above limit $\theta_{0.95}$ in the absolute form:

$$\theta_{0.95} = \pm 2.9 \times 10^{-3} \times 10^{-2} \times 12.6 = \pm 37 \times 10^{-5}\,\mathrm{V}.$$

As this is an accurate measurement, the error of the result is expressed by two significant figures (see Sect. 1.8), and there are no extra figures in the obtained result to be rounded off. The final result is (omitting alternative representations from now on) as follows:

$$U = (12.56368 \pm 0.00037)\ \mathrm{V}\ (0.95)\,.$$

If the measurement was performed with universal estimation of the errors, then the errors of all components would have to be set equal to $5 \times 10^{-3}\%$ and the limit of the measurement error would be

$$\theta'_{0.95} = 1.1 \times 10^{-3}\sqrt{3 \times 5^2} = 0.01\%.$$

Then, in absolute form, $\theta'_{0.95} = \pm 0.0013\,\mathrm{V}$ and the result of measurement would have to be written with fewer significant figures:

$$U = (12.5637 \pm 0.0013)\ \mathrm{V}\ (0.95)\,.$$

Here, two significant figures are retained in the numerical value of the measurement error because the value of its most significant digit is less than 3 (see Sect. 1.8).

8.3 Comparison of Mass Measures

Let us consider the calibration of a 1-kg mass measure by comparing it with the reference standard measure of mass with the same nominal value using a balance. Assume that the comparison was repeated ten times. Column 1 of Table 8.1 lists the measurement results obtained from the comparison of the measures. Our goal is to produce the final measurement result and estimate its inaccuracy.

Assume that the measurement was performed by the methods of precise weighing, which eliminated the error caused by the arms of the balance not having precisely equal length. Thus, it can be assumed that there are no systematic errors.

Table 8.1 presents the input and intermediate data involved in producing the final measurement result and estimating its inaccuracy. Since the systematic errors were eliminated, the measurement results in column 1 can be viewed to be random independent quantities $\{x_i\}$, $i = 1, \ldots, n$ and $n = 10$, and therefore, the probability of all x_i is the same and equal to $1/n$. To simplify the computations, column 2 presents only the varying last three digits of x_i, denoted as x_{i0}.

Their mean value is

$$\bar{x}_{i0} = \frac{1}{n}\sum_{i=1}^{n} x_{i0} = \frac{1}{10} \cdot 7210 \times 10^{-6} = 721 \times 10^{-6}\text{g}.$$

Thus, the estimate of the value of the mass is

$$\bar{x} = 999.998000 + \bar{x}_{i0} = 999.998721 \text{ g}.$$

Table 8.1 Input measurement data and intermediate processing steps in the measurement of the mass of a weight

x_i g	$x_{i0} \times 10^{-6}$ g	$x_{i0} - \bar{x}_{i0} \times 10^{-6}$ g	$(x_{i0} - \bar{x}_{i0})^2 \times 10^{-12}$ g^2
999.998738	738	+17	289
999.998699	699	−22	484
999.998700	700	−21	441
999.998743	743	+22	484
999.998724	724	+3	9
999.998737	737	+16	256
999.998715	715	−6	36
999.998738	738	+17	289
999.998703	703	−18	324
999.998713	713	−8	64
Sum	7,210	0	2,676

We can now obtain the estimate of the variance:

$$S^2(x_i) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i0} - \bar{x}_0)^2.$$

Hence, the standard deviation is

$$S\,(x_i) = \sqrt{\frac{2676}{9} \times 10^{-12}} = 17 \times 10^{-6}\,\mathrm{g}.$$

An estimate of the standard deviation of the obtained value of the mass measure is

$$S_{\bar{x}} = \frac{17 \times 10^{-6}}{\sqrt{10}} = 5 \times 10^{-6}\,\mathrm{g}.$$

We shall find the uncertainty of the result using Student's distribution for confidence probability $\alpha = 0.95$; then, from Table A.2, we find the coefficient t_q for the degree of freedom $\nu = 10 - 1 = 9$ and $q = 1 - \alpha = 0.05$: $t_{0.05} = 2.26$. In accordance with formula (3.20), we obtain the uncertainty of measurement result:

$$u_{0.95} = 2.26 \times 5 \times 10^{-6} = 11 \times 10^{-6}\,\mathrm{g}.$$

Thus, with the confidence probability $\alpha = 0.95$, the mass m of the measure studied lies in the interval

$$999.998\,710\,\mathrm{g} \le m \le 99.998\,732\,\mathrm{g}.$$

The result obtained can be written more compactly as

$$m_{0.95} = (999.998\,721 \pm 11 \times 10^{-6})\,\mathrm{g}.$$

Note that if the data above were processed by the nonparametric methods, the estimate of the measurand would be practically the same but its uncertainty would be much wider (see Sect. 3.8).

8.4 Measurement of Electric Power at High Frequency

As an example of a single independent indirect measurement, consider the measurement of the power generated by a high-frequency current in a resistor. The measurement utilizes the formula $P = I^2R$, where P is the power measured, I is the effective current, and R is the active resistance of the resistor. Measurements of the current and resistance give estimates of their values $\tilde{I}$ and $\tilde{R}$ along with the limits of the relative errors $\delta I = 0.5\%$ and $\delta R = 1\%$.

The errors of measurements of arguments are given in the relative form. Therefore, the influence coefficients are $w'_I = 2$ and $w'_R = 1$. Since the limits of errors of the arguments are known, they can be combined to obtain the uncertainty of the overall measurement result according to (5.49):

$$u_{0.95} = 1.1\sqrt{w_I^2\,(\delta I)^2 + w_R^2\,(\delta R)^2} = 1.1\sqrt{4 \times 0.25 + 1.} = 1.5\%.$$

8.5 An Indirect Measurement of the Electrical Resistance of a Resistor

Consider the measurement of electrical resistance using an ammeter and a voltmeter. This is an indirect measurement with measurement equation $R = U/I$, where R is the electrical resistance of the resistor, U is the voltage drop on the resistor, and I is the strength of the current. Furthermore, it is a dependent indirect measurement because the value of I depends on the value of U.

The connections of the instruments and the resistor are shown in Fig. 8.1. Assume that the measurement was performed under reference conditions for the instruments, and that the input resistance of the voltmeter is so high that its influence on the accuracy of the measurement can be neglected.

The results of measurements of the strength of current and voltage are given in Table 8.2. In accordance with the discussion from Sect. 5.2, all results presented in the table were obtained in pairs: the results with the same subscript belong to the same measurement vector.

We can use in this example both the traditional method and the method of reduction. Let us use each in turn and compare the calculations and results.

8.5.1 Application of the Traditional Method

The traditional method of experimental data processing for dependent indirect measurements was described in Sect. 5.3.

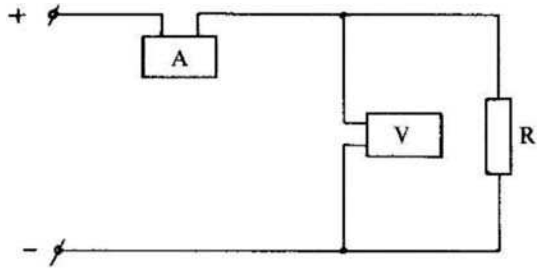


Fig. 8.1 The schema for indirect measurement of an electrical resistance

Table 8.2 Input measurement data in indirect measurement of a resistor

Num.	I_i (A)	U_i (V)
1	0.05996	6.003
2	0.06001	6.001
3	0.05998	5998
4	0.06003	6.001
5	0.06001	5.997
6	0.05998	5.999
7	0.06003	6.004
8	0.005995	5.997
9	0.06002	6.001
10	0.06001	6.003
11	0.05999	5.998

Table 8.3 Data processing for indirect measurement of electrical resistance using the traditional method

Num.	I_i A	U_i V	$(I_i - \bar{I})$ $\times 10^{-5}$ A	$(I_i - \bar{I})^2$ $\times 10^{-10}$ A^2	$(U_i - \bar{U})$ $\times 10^{-3}$ V^3	$(U_i - \bar{U})^2$ $\times 10^{-6}$ V^2	$(I_i - \bar{I})(U_i - \bar{U})$ $\times 10^{-8}$ AV
1	2	3	4	5	6	7	8
1	0.05996	6.003	−3.7	13.69	+2.82	7.95	−10.4
2	0.06001	6.001	+1.3	1.69	+0.82	0.67	+1.1
3	0.05998	5.998	−1.7	2.89	−2.18	4.75	+3.7
4	0.06003	6.001	+3.3	10.89	+0.82	0.67	+2.7
5	0.06001	5.997	+1.3	1.69	−3.18	10.11	−4.1
6	0.05998	5.999	−1.7	2.89	−1.18	1.39	+2.0
7	0.06003	6.004	+3.3	10.89	+3.82	14.59	+12.6
8	0.05995	5.997	−4.7	22.09	−3.18	10.11	+14.9
9	0.06002	6.001	+2.3	5.29	+0.82	0.67	+1.9
10	0.06001	6.003	+1.3	1.69	+2.82	7.95	+3.7
11	0.05999	5.998	−0.7	0.49	−2.18	4.75	+1.5
Sum	0.65997	66.002		74.19		63.61	+29.6

The calculations are illustrated by Table 8.3, which also repeats the input measurement data for convenience. Using the values of U_i and I_i, we obtain the estimates of the arguments:

$$\bar{U} = 66.002/11 = 6.00018\text{V}, \quad \bar{I} = 0.65997/11 = 0.059997\text{A}.$$

We can now compute the estimate of the measurand R. But because the number of measurements of the arguments is the same, one can avoid the inaccuracy of calculation of the argument estimates by obtaining R from the sums of the individual measurement results of the arguments (given in columns 2 and 3, the last row of Table 8.3) rather than from their estimates:

$$\tilde{R} = \bar{U}/\bar{I} = \frac{\sum_{i=1}^{n} U_i}{\sum_{i=1}^{n} I_i} = 66.002/0.65997 = 100.0075\Omega.$$

Now we must calculate the variance and the standard deviation of this result.

First, we will estimate the variances of $\bar{I}$, $\bar{U}$, their standard deviations, and the correlation coefficient. According to the discussion in Sect. 5.2, we obtain

$$S^2(\bar{I}) = \frac{\sum_{i=1}^{n}(I_i - \bar{I})^2}{n(n-1)} = \frac{74.19 \times 10^{-10}}{11 \times 10} = 0.674 \times 10^{-10}\,\mathrm{A}^2,$$

$$S^2(\bar{U}) = \frac{\sum_{i=1}^{n}(U_i - \bar{U})^2}{n(n-1)} = \frac{63.61 \times 10^{-6}}{11 \times 10} = 0.578 \times 10^{-6}\,\mathrm{V}^2.$$

The estimates of standard deviations are

$$S\,(\bar{I}) = 0.82 \times 10^{-5}\mathrm{A},\, S\,(\bar{U}) = 0.76 \times 10^{-3}\,\mathrm{V}.$$

The estimate of the correlation coefficient is

$$r_{I,U} = \frac{\sum_{i=1}^{n}(I_i - \bar{I})}{n(n-1)S(I)S(U)} = \frac{29.6 \times 10^{-8}}{110 \times 0.82 \times 10^{-5} \times 0.76 \times 10^{-3}} = 0.43.$$

It is interesting to note that this correlation coefficient value is statistically insignificant. Indeed, applying a method described in [22], we can check the hypothesis H_0: $\rho_{I,U} = 0$ against H_1: $\rho_{I,U} \neq 0$. The degree of freedom here is $\nu = 11 - 2 = 9$, and we will take the significance level to be $q = 0.05$ as usual, which gives the critical values $t_q = 2.26$ and $r_q = t_q \Big/ \sqrt{t_q^2 + \nu} = 0.60$. Because $0.4 < 0.60$, we must accept H_0 and conclude that the obtained value $r_{I,U} = 0.43$ is not significant, which means that, when the number of measurements n increases, the estimation $r_{I,U}$ of the correlation coefficient will in general decrease. However, it does not mean that the value of $r_{I,U}$ obtained for a *specific sample* can be neglected. On the contrary, it must be always taken into consideration when calculating the estimation of variance for that sample.

In our example, inserting the obtained values into (5.16) we can calculate the desired estimation of standard deviation $S\left(\tilde{R}\right)$. But first we have to calculate the influence coefficients. Thus, the calculations are

$$w_1 = \frac{\partial R}{\partial U} = \frac{1}{I}, \qquad w_2 = \frac{\partial R}{\partial I} = -\frac{U}{I^2},$$

$$\begin{aligned}
S^2(\tilde{R}) &= \left(\frac{\bar{U}}{\bar{I}^2}\right)^2 \times S^2(\bar{I}) + \frac{1}{\bar{I}^2} \times S^2(\bar{U}) - r_{I,U}\frac{\bar{U}}{I^2} \times \frac{1}{I} \times S(\bar{I})S(\bar{U}) \\
&= \left(\frac{6}{36 \times 10^{-4}}\right)^2 \times 0.674 \times 10^{-10} + \frac{1}{36 \times 10^{-4}} \times 0.578 \times 10^{-6} \\
&\quad -2 \times 0.43 \times \frac{6}{36 \times 10^{-4}} \times \frac{1}{6 \times 10^{-2}} \times 0.82 \times 10^{-5} \times 0.76 \times 10^{-3} \\
&= 1.87 \times 10^{-4} + 1.61 \times 10^{-4} - 1.49 \times 10^{-4} \\
&= 1.99 \times 10^{-4}\,\Omega^2,
\end{aligned}$$

and

$$S(\bar{R}) = \sqrt{S^2(R)} = 1.41 \times 10^{-2}\,\Omega.$$

The next step is to find the uncertainty of the obtained result. Unfortunately, we have the standard deviation, but no information about the distribution function of the measurement error, and it is unclear how to find the degree of freedom of the measurement result to account for the dependency between the arguments. Thus, with dependent indirect measurements, we have to use standard deviation of the measurement result as the indication of measurement accuracy rather than its uncertainty (see Sects. 5.3–5.4 for more discussion on the traditional method).

8.5.2 Application of the Method of Reduction

We now turn to the method of reduction described in Sect. 5.5. Table 8.4 lists the intermediate data involved in the calculations. The initial data are again provided in columns 2 and 3.

According to the method of reduction, we first compute values of the measurand using the measurement equation for each measurement vector. The calculated values of R_i $(i = 1, \ldots, 11)$ are given in column 4. Treating these values as if they were obtained by direct measurements, we obtain immediately the estimate of R as

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = 100.0075\Omega$$

Table 8.4 Data processing for indirect measurement of electrical resistance using the method of reduction

Num.	I_i A	U_i V	R_i Ω	$(R_i - \bar{R})$ Ω	$(R_i - \bar{R})^2$ $\times 10^{-2}\,\Omega^2$
1	2	3	4	5	6
1	0.05996	6.003	100.117	+0.109	1.188
2	0.06001	6.001	100.000	−0.002	0.000
3	0.05998	5.998	100.000	−0.002	0.000
4	0.06003	6.001	99.967	−0.041	0.168
5	0.06001	5.997	99.933	−0.075	0.562
6	0.05998	5.999	100.017	+0.009	0.008
7	0.06003	6.004	100.017	+0.009	0.008
8	0.05995	5.997	100.033	+0.025	0.0625
9	0.06002	6.001	99.983	−0.025	0.0625
10	0.06001	6.003	100.033	+0.025	0.0625
11	0.05999	5.998	99.983	−0.025	0.0625
Sum			1,100.083		2.184

and the estimates of its variance and standard deviation as

$$S^2(\bar{R}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(R_i - \bar{R})^2 = \frac{2.184 \times 10^{-2}}{11 \times 10} = 1.99 \times 10^{-4}\,\Omega^2,$$
$$S(\bar{R}) = 1.41 \times 10^{-2}\,\Omega.$$

As one can see from this example, the calculations using the method of reduction are much simpler than using the traditional method, even in this case with a simple measurement equation having only two arguments. More importantly, we now have a set of output data $\{R_i\}$ that does not differ in any way from data obtained in direct measurements. Thus, we know the degree of freedom $\nu = 11 - 1 = 10$ and can compute the uncertainty of the measurement result. Using confidence probability $\alpha = 0.95$ we find the corresponding value of Student's coefficient $t_q = 2.23$ and uncertainty

$$u_{0.95} = 2.23 \times 1.41 \times 10^{-2} = 3.1 \times 10^{-2}\,\Omega.$$

8.6 Measurement of the Density of a Solid Body

The accurate measurement of the density of a solid body can serve as an example of a multiple nonlinear independent indirect measurement. The density of a solid body is given by the formula

$$\rho = m/V,$$

where m is the mass of the body and V is the volume of the body. In the experiment considered, the mass of the body was measured by methods of precise weighing using a balance and a collection of standard weights whose errors did not exceed 0.01 mg. The volume of the body was determined by the method of hydrostatic weighing using the same set of weights. The results of measurements are presented in Table 8.5 in columns 2 and 5.

The difference between the observational results of the body mass is explained by the random error of the balance and the inevitable fluctuations of the environmental conditions. As follows from the data presented, this error is so much larger than the systematic errors in the masses of the weights that the latter errors can be neglected.

8.6.1 Application of the Traditional Method

As the mass of the solid body and its volume are constants, to estimate the density of the body, the mass and volume of the body must be estimated with the required accuracy and their ratio must be formed. For this reason, we find the average values of the measurement results of the arguments and estimates of the standard deviations of these averages (Table 8.5 lists intermediate results for these calculations – the

Table 8.5 Data processing for measurement of the density of a solid body

Num.	Body mass, $m_i \times 10^{-3}$ kg	$(m_i - \bar{m})$ $\times 10^{-7}$ kg	$(m_i - \bar{m})^2$ $\times 10^{-14}$ kg^2	Body volume, $V_i \times 10^{-6}$ m^3	$(V_i - \bar{V})$ $\times 10^{-10}$ m^3	$(V_i - \bar{V})^2$ $\times 10^{-20}$ m^6
1	2	3	4	5	6	7
1	252.9119	−1	1	195.3799	+1	1
2	252.9133	+13	169	195.3830	+32	1,024
3	252.9151	+31	961	195.3790	−8	64
4	252.9130	+10	100	195.3819	+21	441
5	252.9109	−11	121	195.3795	−3	9
6	252.9094	−26	676	195.3788	−10	100
7	252.9113	−7	49	195.3792	−6	36
8	252.9115	−5	25	195.3794	−4	16
9	252.9119	−1	1	195.3791	−7	49
10	252.9115	−5	25	195.3791	−7	49
11	252.9118	−2	4	195.3794	−4	16
Sum			2,132			1,805

deviations of individual measurements from their mean as well as the squares of these deviations):

$$\bar{m} = 252.9120 \times 10^{-3}\,\text{kg}, \qquad \bar{V} = 195.3798 \times 10^{-6}\,\text{m}^3,$$

$$S^2(\bar{m}) = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} (m_i - \bar{m}) = \frac{2132 \times 10^{-14}}{11 \cdot 10} = 19.38 \times 10^{-14}\text{kg}^2,$$

$$S^2(\bar{V}) = \frac{1}{n_2(n_2-1)} \sum_{i=1}^{n_2} (V_i - \bar{V}) = \frac{1805 \times 10^{-20}}{11 \cdot 10} = 16.41 \times 10^{-20}\text{m}^6.$$

The standard deviations of the measurement results of the arguments in the relative form are as follows:

$$S_r(\bar{m}) = \frac{\sqrt{19.38 \times 10^{-14}}}{252.9 \times 10^{-3}} = 1.74 \times 10^{-6},$$

$$S_r(\bar{V}) = \frac{\sqrt{16.41 \times 10^{-20}}}{195.4 \times 10^{-6}} = 2.08 \times 10^{-6}.$$

We can now find the uncertainty of the obtained estimates of the arguments. Both were measured 11 times. Therefore, their degree of freedom is $\nu = 10$. Exploiting the robustness of Student's distribution, we will make use of this distribution. We thus obtain, for confidence probability $\alpha = 0.95$ and the corresponding value of Student's coefficient $t_q = 2.23$, the following confidence limits in relative form:

$$u_{r,0.95}(\bar{m}) = 2.23 \times 1.74 \times 10^{-6} = 3.88 \times 10^{-6},$$
$$u_{r,0.95}(\bar{V}) = 2.23 \times 2.08 \times 10^{-6} = 4.64 \times 10^{-6}.$$

The estimate of the measurand is

$$\bar{\rho} = \frac{\bar{m}}{\bar{\bar{V}}} = \frac{252.9120 \times 10^{-3}}{195.3798 \times 10^{-6}} = 1.294\,463 \times 10^{3}\,\mathrm{kg/m^{3}}.$$

To calculate the uncertainty of the overall measurement result we use the usual method of linearization. It is not difficult to see that, in our example, using just the first term from Taylor's series is sufficient. (To this end, one must estimate the remainder R2 of Tailor's series according to (5.11); we omit these details here.) Thus, we can calculate the standard deviation of the measurement result using formula (5.18). The influence coefficients are $+1$ for mass and (-1) for volume. Hence, the standard deviation of the result in the relative form is as follows:

$$S_r\left(\overline{\rho}\right) = \sqrt{S_r^2\left(\overline{m}\right) + S_r^2\left(\overline{V}\right)} = 2.7 \times 10^{-6}.$$

We shall now find the uncertainty of the result. This can be done in two ways: using the square-root sum formula (5.20) or by taking advantage of the fact that due to the expansion into Taylor's series, the measurement error of the result took the form of a linear combination of the measurement errors of the arguments, making it possible to compute the effective degree of freedom.

In the first method, according to (5.20), the combined uncertainty in relative form is as follows:

$$u_r\left(\overline{\rho}\right) = \sqrt{u_r^2\left(\overline{m}\right) + u_r^2\left(\overline{V}\right)} = \sqrt{3.88^2 + 4.64^2} \times 10^{-6} = 6.0 \times 10^{-6}.$$

We can now apply the correction coefficient W_t introduced in Sect. 5.4, which in our example is $W_t = 0.93$, to arrive at the final result for the combined uncertainty:

$$u_{r,0.95}\left(\overline{\rho}\right) = 0.93 \times 6.0 \times 10^{-6} = 5.8 \times 10^{-6} = 5.8 \times 10^{-4}\%.$$

Using the second method, the effective degree of freedom is given by formula (5.19). The influence coefficients present in that formula are equal to 1 in our case because the measurement errors are represented in relative form. We have already obtained the values of all terms in this formula. Note that in our case $n_1 = n_2 = 11$ and $\nu_1 = \nu_2 = 10$. Thus,

$$\nu_{\mathrm{eff}} = \frac{\left(S^2\left(\overline{m}\right) + S^2\left(\overline{V}\right)\right)^2}{S^4\left(\overline{m}\right)/\nu_1 + S^2\left(\overline{V}\right)/\nu_2} = \frac{(3.03 + 4.33)^2 \times 10^{-24}}{\left(\dfrac{3.03^2}{10} + \dfrac{4.33^2}{10}\right) \times 10^{-24}} = 19.7.$$

For $\nu_{\mathrm{eff}} = 19.7$ and $\alpha = 0.95$ we find from the table of Student's distribution (see Appendix A2) that $t_q = 2.09$. From here, the combined uncertainty $u_r'\left(\bar{\rho}\right)$ (so denoted to distinguish from the uncertainty computed according to the first method) becomes as follows:

$$u'_{r,0.95}(\bar{\rho}) = 2.09 \times 2.7 \times 10^{-6} = 5.6 \times 10^{-4}\%.$$

As we can see, both methods produce practically the same results. This is natural because both methods are based on the same assumption that the measurement errors of the arguments are normally distributed. In applying the square-root sum formula this assumption allows one to compute the correction factor, in using the effective degree of freedom, to construct the composition of the distributions of the measurement errors of the arguments. If in fact these distributions significantly deviate from normal, these deviations would lead to the same inaccuracy of the obtained uncertainty in both methods. Thus, the matching results under the two methods in themselves do not prove their accuracy.

8.6.2 Application of the Method of Transformation

We start with the same measurement data that were used in the traditional method and that are presented in Table 8.5 in Sect. 8.6.1. The measurement equation remains as before:

$$\rho = \frac{m}{V}.$$

This is a multiplicative measurement equation $\rho = f_V(V) \cdot f_m(m)$, where $f_V(V) = 1/V$ and $f_m(m) = m$. Let us first substitute argument V by its estimate $\bar{V}$, retaining $m$as a deriving argument. The transformation coefficient for m is $C_m = \Psi_m(V) = 1/V$ and its estimate is $1/\bar{V}$. Then, for each value m_i of argument m, (5.27) gives $\rho_{m,i} = \frac{m_i}{\bar{V}}$. The mean value $\bar{V}$ was calculated earlier when we applied the traditional method – its value is $\bar{V} = 195.3798 \times 10^{-6}\,\mathrm{m}^3$. The elements of the output group $\rho_{m,i}$ produced by the deriving argument values m_i are shown in column 2 of Table 8.6.

Now we substitute m with its estimate $\bar{m}$ and take V as a deriving argument, thus obtaining $\rho_{V,i} = \frac{\bar{m}}{V_i}$ for each value V_i. Here, $C_V = \Psi_V(m) = m$, and its estimate is $\bar{m}$. This mean value $\bar{m} = 252.9120 \times 10^{-3}$ kg was calculated above in Sect. 8.6.1. The output group $\rho_{V,i}$ is shown in column 2 of Table 8.7.

We thus have two output groups of data of the measurand, each derived from the group of measurement data of the corresponding argument. Using the procedure of Sect. 5.6, we can now obtain the estimate of the density and its uncertainty.

The estimate of the measurand can be found from (5.32), although in our case both output groups have the same means and so the overall mean will be the same:

$$\bar{\bar{\rho}} = 1.2944629 \times 10^3\,\mathrm{kg/m}^3.$$

Table 8.6 The output group derived from argument m (the mass of a body)

I	$\rho_{m,i} \times 10^3\ \mathrm{kg/m^3}$	$(\rho_{m,i} - \bar{\rho}_m)$ $\times 10^{-3}\ \mathrm{kg/m^3}$	$(\rho_{m,i} - \bar{\rho}_m)^2$ $\times 10^{-6}\ (\mathrm{kg/m^3})^2$
1	1.2944626	−0.3	0.1
2	1.2944698	6.9	47.6
3	1.2944790	16.1	259.2
4	1.2944682	5.3	28.1
5	1.2944575	−5.4	29.2
6	1.2944498	−13.1	171.6
7	1.2944595	−3.4	11.6
8	1.2944605	−2.4	5.8
9	1.2944626	−0.3	0.1
10	1.2944605	−2.4	5.8
11	1.2944621	−0.8	0.6
Average	1.2944629		Σ 559.6

Table 8.7 The output group derived from argument V (the volume of a body)

I	$\rho_{\mathrm{v},i} \times 10^3\ \mathrm{kg/m^3}$	$(\rho_{m,i} - \bar{\rho}_m)$ $\times 10^{-3}\ \mathrm{kg/m^3}$	$(\rho_{m,i} - \bar{\rho}_m)^2$ $\times 10^{-6}\ (\mathrm{kg/m^3})^2$
1	1.2944626	−0.3	0.1
2	1.2944420	−20.9	436.8
3	1.2944685	5.6	31.4
4	1.2944493	−13.6	185.0
5	1.2944652	2.3	5.3
6	1.2944698	6.9	47.6
7	1.2944672	4.3	18.5
8	1.2944659	3.0	9.0
9	1.2944679	4.9	24.0
10	1.2944679	4.9	24.0
11	1.2944659	3.0	9.0
Average	1.2944629		Σ 790.7

Let us now turn to the estimation of uncertainty of the obtained result. We begin with finding parameters of the error distribution. The variance and standard deviation of the random error of the above mean in relative form can be computed according to (5.33), having in mind that $n_m = n_V = n$ and $Z = n_m + n_V = 2n = 22$. Thus,

$$S^2_{\Psi,rel}\left(\overline{\overline{\rho}}\right) = \frac{1}{\overline{\overline{\rho}}^2} \frac{\sum_{i=1}^{n}\left(\rho_{m,i} - \overline{\overline{\rho}}\right)^2 + \sum_{i=1}^{n}\left(\rho_{V,i} - \overline{\overline{\rho}}\right)^2}{2n\,(2n-1)} = \frac{559.6 + 790.7}{1.294 \times 10^3 \times 22 \times 21}$$
$$= 1.74 \times 10^{-12}.$$

The standard deviation therefore is

$$S_{\Psi,rel}\left(\bar{\rho}\right) = 1.32 \times 10^{-6}.$$

The conditionally constant errors must be estimated separately for each group. Let us start with the group derived from argument m, the mass of the body. Given that $\tilde{C}_m = 1/\bar{V}$ and V's influence coefficient is $w_V = \left(\frac{d\Psi_m}{dV}\right)_{V=\bar{V}}$, (5.31) becomes as follows:

$$\varepsilon\left(x_{m,i}\right) = \varepsilon\left(\tilde{C}_m\right) = \frac{w_V}{\tilde{C}_m}\delta\left(\bar{V}\right) = \frac{\delta\left(\bar{V}\right)}{\bar{V}}.$$

From here, we obtain

$$S^2_{\vartheta,rel}\left(\bar{\rho}_m\right) = \frac{S^2\left(\bar{V}\right)}{\bar{V}^2} = S^2_{rel}\left(\bar{V}\right).$$

We already computed $\bar{V} = 195.3798 \times 10^{-6}\,\mathrm{m}^3$ and $S_{rel}\left(\bar{V}\right) = 2.08 \times 10^{-6}$ in Sect. 8.6.1, and we can now compute the confidence interval for the conditionally constant error in the output group we are considering. The group has 11 observations, so the degree of freedom is $\nu = 10$. For confidence probability $\alpha = 0.95$, the quantile of Student's distribution is $t_{m,\vartheta} = 2.23$. Thus, the members of the output group $\{\rho_{m,i}\}$ and its mean all have the following confidence limit of the conditionally constant error, in relative form:

$$\theta_{m,rel} = t_{m,\vartheta} S_{\vartheta,rel}\left(\bar{\rho}_m\right) = 2.23 \times 2.08 \times 10^{-6} = 4.64 \times 10^{-6}.$$

We repeat the same calculations for the other output group, which has been derived from V. For this group, $\tilde{C}_V = \bar{m}$, $\Psi_V\left(m\right) = m$, and $w_m = \frac{d\Psi_V}{dm} = 1$. Therefore, we have

$$S^2_{\vartheta,rel}\left(\bar{\rho}_V\right) = S^2_{rel}\left(\bar{m}\right).$$

In Sect. 8.6.1, we computed $\bar{m} = 252.9120 \times 10^{-3}\,\mathrm{kg}$ and $S^2_{rel}\left(\bar{m}\right) = 3.02 \times 10^{-12}$. Thus, $S_{rel}\left(\bar{\rho}_V\right) = S_{rel}\left(\bar{m}\right) = 1.74 \times 10^{-6}$. We again have the degree of freedom $\nu = 10$ and we must use the same confidence probability $\alpha = 0.95$, hence the quantile of Student's distribution will be the same also: $t_{V,\vartheta} = t_{m,\vartheta} = t_\vartheta = 2.23$. The confidence limit of the conditionally constant error for this group is therefore

$$\theta_{V,rel} = t_{V,\vartheta} S_{\vartheta,rel}(\bar{\rho}_V) = 2.23 \times 1.74 \times 10^{-6} = 3.88 \times 10^{-6}.$$

The obtained limits of the systematic errors differ for each output group. Thus, the overall limit of the systematic error must be calculated as a weighted mean, which reduces to the simple mean in our case because both groups have the same number of elements. Hence, the overall confidence limit of the systematic error is

$$\theta_{rel}(\bar{\bar{\rho}}) = \frac{\theta_{m,rel} + \theta_{V,rel}}{2} = 4.26 \times 10^{-6}.$$

The relative standard deviation of the overall conditionally constant error is also equal to the mean of the relative standard deviations obtained from individual output groups:

$$S_{\vartheta,rel}(\bar{\bar{\rho}}) = \frac{S_{\vartheta,rel}(\bar{\rho}_m) + S_{\vartheta,rel}(\bar{\rho}_V)}{2} = \frac{2.08 + 1.74}{2} \times 10^{-6} = 1.91 \times 10^{-6}.$$

Correspondingly, we also obtain the relative variance of this error:

$$S^2_{\vartheta,rel} = 3.65 \times 10^{-12}.$$

The combined relative standard deviation of the obtained measurement result is

$$S_c = \sqrt{S^2_{\vartheta,rel} + S^2_{\Psi,rel}} = \sqrt{(3,65 + 1.74) \times 10^{-12}} = 2.32 \times 10^{-6}.$$

Following now the procedure for combining errors, we compute coefficients t_q and t_ϑ, both for confidence probability $\alpha = 0.95$. For this confidence probability and $\nu = 10$, we already have $t_\vartheta = 2.23$. For the random component, the degree of freedom is $\nu = Z - 1 = 21$ (see Sect. 5.6), and coefficient $t_q = 2.08$. Thus, the combined coefficient t_c is as follows:

$$t_c = \frac{t_\vartheta S_{\vartheta,rel} + t_q S_{\Psi,rel}}{S_{\vartheta,rel} + S_{\Psi,rel}} = \frac{2.23 \times 1.91 + 2.08 \times 1.32}{1.91 + 1.32} = 2.17.$$

Thus, the combined uncertainty becomes as follows:

$$u_{c,rel} = t_c S_c = 2.17 \times 2.32 \times 10^{-6} = 5.0 \times 10^{-6} = 5.0 \times 10^{-4}\%.$$

We now have the measurement result in its final form:

$$\rho_{0.95} = (1.294463 \times 10^3 \pm 0.65)\,\text{kg/m}^3.$$

As we can see, the estimate of the measurand obtained here is exactly the same as with the traditional method; this is natural because both methods use the same mean values of argument observations. However, the uncertainty of the result in the traditional method was 6.0% when computed with the square-root sum formula, 5.8% in the same method after applying the correction, and 5.6% with the effective degree of freedom, while the method of transformation produced 5.0%. The latter result appears more justified for the following reasons:

1. The formula for computing effective degree of freedom is approximate and is obtained based on the assumption that errors are normally distributed.
2. The square-root sum formula is correct for summation of variances and not uncertainties. The analysis we gave in Chap. 5 showed that using this formula for summation of uncertainties leads to exaggerated result even if the initial errors are normally distributed.

In contrast, the method of transformation in the case of two arguments does not rely on assumptions for which it is unclear whether or not they hold in the measurement at hand.

8.7 Measurement of Ionization Current

Accurate measurements of weak currents, for example, currents generated by γ rays from measurement standards of unit radium mass are performed by the compensation method using an electrometer. Such currents are measured and compared, for example, in the process of calibration of these standards.

In the compensation method, a high-impedance resistor is inserted into the circuit with the current to be measured. This resistor is also connected in parallel to a capacitor, which is charged prior to being connected. The two connections are arranged so that the measured current and the discharge current from the capacitor flow in the opposite directions. The difference between the two currents creates voltage on the resistor, which is detected by the electrometer. When the electrometer indicator shows zero, the two currents are equal. The time from the start of the capacitor's discharge to when the two currents equalize is measured; this time depends on the dynamics of the capacitor discharge, which is determined by the time constant of the circuit containing the capacitor and resistor. This constant can be determined accurately because both the capacitance of the capacitor and the impedance of the resistor are found a priori with high accuracy. Thus, given the known charge on the capacitor before it is connected to the resistor, one can determine the ionization current by the discharge time until the moment of compensation.

The measured strength of current I is defined by the expression

$$I = CU/\tau,$$

where C is the capacitance of the capacitor used to compensate the ionization current; U is the initial voltage on the capacitor; and τ is the compensation time. As U and τ are dependent, it is a dependent measurement.

We shall examine the measurement of ionization current on the specific apparatus described in [33]. It employs a capacitor with capacitance $C = 4006.3$ pF, which is known to be within 0.005% of the above value. The voltage on the capacitor is established with the help of a class 0.1 voltmeter with a measurement range of $0 - 15$ V. The time is measured with a timer whose scale is divided into tenths of a second. The results of a calibration of one standard of radium mass against another using this apparatus are presented in [34]; we will use these results to estimate the accuracy of the measurement of the ionization current involved in the calibration procedure.

The measurement described in [34] included 27 repeated observations. Each time the same indication of the voltmeter $U = 7$ V was established and the compensation time was measured. The results of the 27 observations of time are given in the first column of Table 8.8. Using the measurement equation, we can compute the strength of the ionization current from the compensation time. The 27 values of the current

Table 8.8 Measurement results and intermediate processing steps in the measurement of ionization current

τ s	$I_i \times 10^{-10}$ A	$(I_i - \bar{I}) \times 10^{-14}$ A	$(I_i - \bar{I})^2 \times 10^{-28}$ A^2
74.4	3.7694	7	49
74.6	3.7593	−94	8,836
74.3	3.7745	58	3,364
74.6	3.7593	−94	8,836
74.4	3.7694	7	49
74.4	3.7694	7	49
74.4	3.7694	7	49
74.4	3.7694	7	49
74.4	3.7694	7	49
74.3	3.7745	58	3,364
74.5	3.7643	−44	1,936
74.4	3.7694	7	49
74.5	3.7643	−44	1,936
74.4	3.7694	7	49
74.6	3.7593	−94	8,836
74.2	3.7705	18	324
74.5	3.7643	−44	1,936
74.3	3.7745	58	3,364
74.4	3.7694	7	49
74.4	3.7694	7	49
74.5	3.7643	−44	1,936
74.5	3.7643	−44	1,936
74.3	3.7745	58	3,364
74.3	3.7745	58	3,364
74.3	3.7745	58	3,364
74.4	3.7694	7	49
74.5	3.7643	−44	1,936

corresponding to the measured compensation times are listed in column 2 of the table. We now need to obtain the estimate of the result of this measurement and its inaccuracy.

Let us first obtain the estimate of the current. Because ionization currents are weak, one has to account for the so-called background current caused by the background radiation. The average background current is usually equal to $(0.5 - 1) \times 10^{-12}$ A and can be measured to within 5%. In the measurement in question, the background current was found to be $\bar{I}_b = 0.75 \times 10^{-12}$ A. The average value of current observations from Table 8.8 is $\bar{I} = 3.7687 \times 10^{-10}$ A. Thus, the estimate the ionization current I_x is

$$\tilde{I}_x = \bar{I} - \bar{I}_b = 3.7612 \times 10^{-10}\ \text{A}.$$

Now let us turn to the inaccuracy. First consider the conditionally constant systematic errors. For a class 0.1 voltmeter, its limit of error in indicating the voltage of 7 V is $\theta_U = 0.1 \times (15/7) = 0.21\%$. The limit of error of measuring compensation time with the timer that has the graduations of 0.1 s is equal to half the graduation or 0.05 s. In relative form, for the time intervals of 74–75 s, this gives $\theta_\tau = (0.05/74) \times 100 = 0.067\%$. Although the capacitance of the capacitor is supposed to be known within 0.005%, the measurement was performed under rated rather than reference temperature conditions, leading to an additional error. Thus, the capacitance is known only with the limit of error of 0.05%. The limit of measurement error of the background current, which is within 0.5% of the value of the background current, is only 0.013% with respect to the ionization current estimate, and it can obviously be neglected compared to the error in voltage indication θ_U. Turning to formula (5.44) and taking confidence probability $\alpha = 0.95$, we obtain

$$\theta_{I,0.95} = k\sqrt{\theta_C^2 + \theta_U^2 + \theta_\tau^2} = 1.1\sqrt{0.05^2 + 0.21^2 + 0.067^2} = 0.24\%.$$

Now let us consider random errors. We will use the method of reduction. First we shall find an estimate of the standard deviation of the measurement result, which is

$$S\left(\bar{I}_x\right) = S\left(\bar{I}\right) = \sqrt{\frac{\sum\limits_{i=1}^{27}\left(I_i - \bar{I}\right)^2}{27 \times 26}} = 9 \times 10^{-14}\,\text{A}.$$

It is obvious that the random error can be neglected compared to the limit of the conditionally constant systematic error computed above, which in the absolute form is equal to $\theta_I = 0.009 \times 10^{-10}$ A. The latter therefore determines the overall inaccuracy of the result. Therefore, our obtained estimate of the ionization current has one extra digit. Rounding it off, we arrive at the result of the measurement:

$$I_x = (3.761 \pm 0.009) \times 10^{-10}\,\text{A}\,(0.95)\,.$$

Finally, as a side note, Table 8.8 shows that the random error of an individual observation in this measurement, which could be explained by the inaccuracy of the detection of the moment of the equality of the measured and compensating currents and of the setting of the initial voltage on the capacitor, can reach 0.25% (this can be seen as the deviation of individual observations in Table 8.8, column 2, from the average). However, repeating the measurement 27 times allowed us to reduce the error to the level where it could be neglected compared to the systematic errors.

8.8 Measurement of the Activity of a Radioactive Source

We shall examine the measurement of the activity of a radioactive source by absolute counting of α particles emitted by the source. We will use the experiment described in [15], as well as the measurement data reported there, as the basis for our discussion. The measurement is performed using a detector that counts the particles arriving from the source through a diaphragm opening. The number of particles captured by the detector depends on the geometric configuration of the experimental setup – the diameter of the diaphragm, the distance between the detector and the source, and the diameter of the source (assuming the source is spherical). Following [15], these parameters can be encapsulated into a geometric factor G, which is calculated from the above quantities. Then the measured radioactivity is determined from the formula

$$A = GN_0\eta,$$

where G is the geometric factor of the apparatus, N_0 is the α-particle counting rate, and η is the α-particle detection efficiency. In the course of the measurement, G does not change, so that errors of G create a systematic error of measurement of the activity A. Measurements of the numbers of α particles, however, have random errors.

To reduce the error arising from the error of the geometric factor, the measurements were performed for different values of this factor (by changing the distance between the source and detector and the diameter of the diaphragm). All measurements were performed using the same source ^{239}Pu.

All the arguments appear in the measurement equation with the same degree of 1. Thus, as discussed in Sect. 5.7, it is convenient to express their errors in relative form since all the influence coefficients will then be equal to 1. Table 8.9 gives measurement results for the five geometric configurations studied. In each case, 50 measurements were performed, and estimates of the measured quantity and their standard deviation, which are also presented in Table 8.9, were calculated. The standard deviations of the (conditionally constant) systematic errors of the results were calculated from the estimated limiting values of all error components under the assumption that they can be regarded as centered uniformly distributed random quantities.

The data in Table 8.9 show, first of all, that the systematic errors are much larger than the random errors, so that the number of measurements in the groups was sufficient. The observed difference between the obtained values of the activity of the nuclides in the groups can be mostly explained by their different systematic errors.

In the example studied, the same quantity was measured in all cases. Therefore, one can use the weighted mean as the overall estimate of the measurand. Based on

Table 8.9 The results of measurements of the activity of nuclides using a setup with different geometric factors

Group number j	Source-detector distance (mm)	Diaphragm radius (mm)	Measurand estimate $x_j \times 10^5$	Estimates of standard deviation	
				Random errors (%)	Systematic errors (%)
1	97.500	20.017	1.65197	0.08	0.52
2	97.500	12.502	1.65316	0.10	0.52
3	397.464	30.008	1.66785	0.16	0.22
4	198.000	20.017	1.66562	0.30	0.42
5	198.000	30.008	1.66014	0.08	0.42

Table 8.10 The estimates of combined variances and weights of measurement results by different geometric factors

Group number j	Estimate of combined variance $S^2(\bar{x}_j)$ (%)^2	Weight g_j
1	0.28	0.12
2	0.28	0.12
3	0.07	0.46
4	0.27	0.12
5	0.18	0.18

the considerations from Sect. 7.5, we shall use (7.13) to calculate the weights. First, we shall calculate an estimate of the combined variance according to (7.12):

$$S^2\left(\bar{x}_j\right) = S_\psi^2\left(\bar{x}_j\right) + S_\vartheta^2\left(\bar{x}_j\right).$$

The results of the calculations are given in Table 8.10. As an example, we provide the calculation details of weight g_1:

$$g1 = \frac{\dfrac{1}{0.28}}{\dfrac{1}{0.28} + \dfrac{1}{0.28} + \dfrac{1}{0.07} + \dfrac{1}{0.27} + \dfrac{1}{0.18}} = \frac{3.57}{30.7} = 0.12$$

Now we find the weighted mean:

$$\tilde{A} = \bar{\bar{x}} = \sum_{j=1}^{5} g_j \bar{x}_j = 1.6625 \times 10^5.$$

We can now estimate the uncertainty of the measurement result. To do this, we need to find, using (7.13), the standard deviations of the random and conditionally constant systematic components of the weighted mean and then, since $S\left(\tilde{A}\right)$ has already been found, calculate t_c from (4.22). All data for these calculations are available in Tables 8.6 and 8.7.

The standard deviations of the random and systematic components of the weighted mean are as follows:

$$S_\psi^2\left(\bar{\bar{x}}\right) = \sum_{j=1}^{L} g_j^2 S_\psi^2\left(\bar{x}_j\right) = 71.58 \times 10^{-8} \text{ and } S_\psi\left(\bar{\bar{x}}\right) = 8.46 \times 10^{-4},$$

$$S_\vartheta^2\left(\bar{\bar{x}}\right) = \sum_{j=1}^{L} g_j^2 S_\vartheta^2\left(\bar{x}_j\right) = 261.7 \times 10^{-8} \text{ and } S_\vartheta\left(\bar{\bar{x}}\right) = 16.2 \times 10^{-4}.$$

Next, we compute the uncertainty of the systematic component, θ_α. The easiest way to do it is by using (4.3). For this, however, we need to transfer from the standard deviations of the elementary systematic errors back to their limits, which as we know can be done using factor $\sqrt{3}$ (since $S^2 = \theta^2/3$). Thus,

$$\theta_\alpha = k\sqrt{3\sum_{j=1}^{L} g_j^2 S_\vartheta^2\left(\bar{x}_j\right)} = k\sqrt{3S_\vartheta^2\left(\bar{\bar{x}}\right)}.$$

Taking $\alpha = 0.95$, we have $k = 1.1$ and $\theta_{0.95} = 1.1 \times 1.73 \times S_\vartheta\left(\bar{\bar{x}}\right) = 1.90 S_\vartheta\left(\bar{\bar{x}}\right)$. From here, we obtain $t_\vartheta = \theta_{0.95}/S_\vartheta\left(\bar{\bar{x}}\right) = 1.90$. To find quantile t_q of Student's distribution for the selected confidence probability, we also need the degree of freedom. In general, when the measurement result represents a weighted mean of several measurements, the degree of freedom is obtained from (5.19) as an effective degree of freedom. In our case, however, we have five groups, each comprising a large number of observations ($n = 50$ in each group), so it is obvious even without calculations that the resulting distribution can be considered normal. Then, $t_q = z_{\frac{1-\alpha}{2}} = 1.96$. We can now use formula (4.22) to find t_c:

$$t_c = \frac{t_\vartheta S_\vartheta\left(\bar{\bar{x}}\right) + t_q S_\psi\left(\bar{\bar{x}}\right)}{S_\vartheta\left(\bar{\bar{x}}\right) + S_\psi\left(\bar{\bar{x}}\right)} = 1.92.$$

Finally, we are ready to compute the uncertainty of the measurement result:

$$u_c = t_c S\left(\bar{\bar{x}}\right) = 1.92 \times 0.182 = 0.35\%.$$

In the form of absolute uncertainty, we obtain $u_{0.95} = 0.006 \times 10^5$. Thus, the result of the measurement can be given as follows:

$$A = (1.662 \pm 0.006) \times 10^5\ (0.95).$$

Chapter 9
Conclusion

9.1 Measurement Data Processing: Past, Present, and Next Steps

Historically, metrology emerged as a science of measures. Even in the middle of the last century, metrology was considered to be the science of measurements concerning the creation and maintenance of measurement standards [37]. With this approach, the theory of accuracy of measurements was limited to the problems of estimation of the accuracy of multiple measurements and only to random errors. Math statistics was a natural fit for these problems. As a result, the science of measurement data processing was in essence the reformulation of math statistics in the context of random errors.

This state of affairs can be clearly seen by examining relatively recent books on the subject, for example, *Data Analysis for Scientists and Engineers* by S. Meyer (1975), *Data Reduction and Error Analysis for Physical Sciences* by Ph. Bevington and D. Robinson (1992), and *Measurement Theory for Engineers* by I. Gertsbakh (2003). Even the book *The Statistical Analysis of Experimental Data* (National Bureau of Standards, 1964) by J. Mandel, which stands out by considering concrete measurement problems, remained within the above confines. Nevertheless, because this purely mathematical theory found practical applications, even in a restricted case of random errors in multiple measurements, this theory obtained the status of the classical theory of measurement data processing.

In the meantime, this theory did not satisfy practical needs. In particular, every practitioner knew that in addition to random errors, there are systematic errors, and the overall inaccuracy of the measurement result combined both of these components. But the classical theory ignored this fact and, furthermore, not so long ago considered it incorrect to combine these two components. There were other practical problems ignored by the classical theory. As a result, those who encountered these problems in their practice resorted to ad hoc and often incorrect methods. For example, in the case of single measurements, the measurement errors were often equated to the fiducial error of the measuring device used (see Chap. 2), which is wrong. To account for systematic errors in a multiple measurement, people often simply added them to the random errors, which overestimated the inaccuracy of the result.

S.G. Rabinovich, *Evaluating Measurement Accuracy: A Practical Approach*, DOI 10.1007/978-1-4419-1456-9_9,

The classical theory further ignored single measurements whereas these measurements are the most commonly used in industry, scientific research, and trade. Yet another limitation concerned the calculation of the inaccuracy of dependent indirect measurements, which are typical in scientific experiments. The classical theory did not offer ways to estimate the inaccuracy of these measurements as a confidence interval, forcing scientists to make do with the standard deviation as the characteristic of inaccuracy of the measurement result. Standard deviation is an indirect and sometime ambiguous characteristic of inaccuracy, while confidence interval is intuitive, unambiguous, and reflects the inaccuracy directly.

Practical needs demanded solutions to these and other problems not handled by the classical theory. In the first edition of the "International Vocabulary of Basic and General Terms in Metrology" published by ISO in 1984, metrology was declared to be the science of measurements, regardless of their accuracy or application field.

To address these problems, a new theory started to take shape toward the end of the last century. This theory does not obviate but subsumes the classical theory and augments it with accounting for physical meaning of the metrological problems being addressed. We therefore can call it the physical theory of measurement data processing.

By considering the physical meaning of metrological problems, the new theory has offered the method of reduction, which allows one to calculate the confidence interval for the result of a dependent indirect measurement. Furthermore, this method removes the need for the correlation coefficient in experimental data processing, leading to a simpler and more accurate calculation procedure.

The new theory has also resulted in a clear and simple method for combining systematic and random errors in a measurement result. The analysis of this method showed its high accuracy. This new theory has also revealed an organic connection between single and multiple measurements and thus introduces into the analysis of inaccuracy of measurements the properties of measuring instruments. Besides providing solutions to these and other specific practical problems, the physical theory also considers the foundational issues of measurements.

This book offers systematic treatment of the physical theory and in this way defines this new discipline. At the same time, this book obviously does not exhaust this subject, and a number of problems still await their solutions. We list some of these gaps below.

- The theory of single measurements requires further development, especially in regard to accounting for the errors of measuring instruments. A complicating factor in this problem is a large variety of measuring instrument types for which suitable techniques must be developed.
- Although the diversity of measuring instruments prohibits the development of the general theory of their design, it is possible and necessary to develop a general theory of accuracy of measuring instruments. The accuracy is the common aspect that unites these devices. This book takes an initial step toward such a theory, but much more work is required.
- A large and common class of measurements involving recording instruments (such as analog or digital automatic plotters, XY-recorders, etc.) came to be

known as dynamic measurements [27, 51]. There are many open problems in dynamic measurements; among them is an attractive problem to find the form and parameters of an input signal having the recorded output signal and knowing the dynamic properties of the recorder. Modern computers make solving this problem feasible.
- The errors and uncertainty of measurements are always estimated in an indirect way, and the calculations include some assumptions. However, the correctness of these assumptions and the validity of the resulting estimates have never been experimentally checked. A general approach to filling this important gap would involve measuring the same measurand in parallel by different methods, with one method being an order of magnitude more accurate than the other, and then comparing the measurement results and the calculated uncertainties. A promising alternative here is to use the Mote Carlo method, employing precisely known analytical expressions to generate input data. By comparing the results of experimental data processing based on certain assumptions (e.g., the assumption that a conditionally constant error is a uniformly distributed random variable) with the results of the Monte Carlo method using simulated data generated to comply with those assumptions, one can estimate the implications of the assumptions made.
- The application of the square-root sum method to uncertainty calculations requires further investigation. In particular, we analyzed this method in the present book for random errors having a normal distribution. An important question is whether this method can be used for other distributions, and how accurate it would be.
- The applicability of the least-squares method to experimental data processing when residuals are not purely random quantities. It is known that the least-squares method is optimal when residuals are normally distributed random quantities. However, residuals can include both systematic and random errors. Although the least-squares method has been considered for random residuals only, it is promising in these more general cases because it naturally accounts for both types of errors. In fact, it is sometimes used in these cases without theoretical justification. However, its behavior in these cases is unknown.

Although this list of problems is subjective and incomplete, it suffices to show that the physical theory of measurement data processing is a live discipline still under development.

9.2 Remarks on the International Vocabulary of Metrology

The first edition of "International Vocabulary of Metrology – Basic and General Concepts and Associated Terms" (VIM) was prepared by Working Group 2 of JCGM and published by ISO/IEC in 2007. The Foreword of VIM states that this document replaces all previously published editions of the International Vocabulary of Basic and General Terms in Metrology. The new VIM differs significantly from the previous one and reconsiders the definitions of many terms. Four terms defined in this VIM, *measurement result*, *true value, error,* and *uncertainty*, have a particular bearing on the present book and are appropriate to be discussed here.

The clause 2.9 of the VIM defines *measurement result* as a "*set of quantity values being attributed to a measurand together with other available relevant information.*" Note 1 clarifies that "…this may be expressed in the form of a probability density function (PDF)."

According to this definition, a set of observations obtained from the multiple measurements represents the result of the measurement. However, as known, the goal of the measurement is always a single estimate for the measured quantity obtained from the analysis of this set, often augmented with an indication of its accuracy, but not the set itself. Having a single estimate allows one to use measurement results in mathematical formulas expressing natural laws. One cannot replace values with distribution functions in these calculations. Therefore, this definition of *measurement result* is not productive, and the traditional definition should be retained, which is that the *measurement result is a value attributed to a measurand, obtained by measurement*.

The definition of *true value* (VIM, clause 2.11) says that it is the "*quantity value consistent with the definition of a quantity.*" However, even the value assigned to the measurand as the result of the measurement is consistent with the definition of the quantity; otherwise, it would be useless. In other words, this definition of the true value suggests that the value of the quantity and its true value are the same. However, the established meaning of the term true value is that it is an abstract, unreachable, property of the measurand. Without this established meaning, it is impossible to define the accuracy of a measurement. Therefore, the following definition, derived from the one in [12], is advisable: *true value – the value of a quantity that being known would ideally reflect the property of an object with respect to the purpose of the measurement.* Note: as any ideal, the true value is impossible to find.

The definition of *true value* in VIM has three notes, two of which require a discussion. Note 1 states: "In the Error Approach … a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity." Two aspects of this note are objectionable.

First, we would like to disagree with the notion that the incomplete amount of detail in the definition of a quantity entails a set of true values rather than a single true value for the quantity. It is well known that the goal of any measurement is to obtain a numeric value that reflects the measured quantity. Measurement results realize this goal. It is this aspect of measurements that allows us to apply mathematics to natural sciences, and it is only possible if every measured quantity has a single true value. Indeed, if we assumed that the measured quantity had multiple true values, it would be impossible to associate it with a single number and use it in subsequent mathematical formulas. Although a measurement result often includes an indication of its accuracy, and this indication is often expressed as an interval, any measurement result still assigns a value (usually taken as the most likely value within the interval) to the measurand.

The concept of the true value of a measured quantity is considered in detail in Sect. 1.4 of the present book. That section also considers the example of the measurement of the thickness of a sheet of a material, which is presented in GUM (Sect. D.3.2 and D.3.4) to motivate the idea of a measured quantity having a set of true values. We explained that when the thickness of the sheet is different in different places and one must reflect these different thickness values by measuring the thickness in different places, we have in fact several distinct measurements, one in each place of the sheet. Each given point of the sheet has its own true value of thickness and will have its own measurement result. There is no single measurement result here, and the set of true values does not have to do with individual measurements of the sheet thickness in different points. Thus, this example does not show the need or the usefulness of having a set of true values for one measured quantity.

Regarding the inherently incomplete amount of detail reflected in the definition of the quantity, the definition of the quantity must only reflect the property that is of interest to the experimenter. The lack of detail in the definition of the quantity is not a reason for introducing a set of true values for the quantity.

Second, we question the usefulness of distinguishing two approaches to estimation of the accuracy of measurements. Defining new approaches is beneficial only if they enable solutions to new problems. However, the VIM does not present any new problem solved by the Uncertainty Approach with its set of true values for a quantity. Thus, its introduction appears unwarranted. Further, the sentence following the note in question mentions additional approaches but leaves it unclear what these approaches are. From the above considerations, we conclude that the notion of a "set of true values" must be removed from VIM.

Note 3 also raises objections. It represents an attempt to justify an erroneous concept of the "Guide to the Expression of Uncertainty in Measurement" [2] of the equivalency between the true value and the value of the measured quantity. However, the true value is an unreachable ideal concept, while the value of a measured quantity is a measurement result. Thus, the two cannot be equivalent no matter the accuracy of the measurement in question. We return to this issue in more detail in Sect. 9.3.

These considerations lead to a conclusion that Notes 1 and 3 should be removed from VIM.

The definition of *measurement error* given in clause 2.16 says: *measured quantity value minus a reference quantity value*. Unfortunately, the above sentence cannot be considered a definition because it does not explain the meaning of the term, and instead it attempts to provide an algorithm for its calculation. As a matter of fact, that algorithm is unrealistic: it follows from clause 5.18 that the reference quantity value in measurements refers to the true value, which is always unknown. Furthermore, this definition narrows the meaning of the term since it only covers the absolute error, leaving a commonly used relative error aside.

I consider *measurement error* to be properly defined as *a deviation of the result of measurement from the true value of the measurand.* This definition is not algorithmic and makes it clear that just like the true value, measurement error is impossible to obtain. In fact, the above consideration warrants the following note to this definition: Because the true value is always unknown, the error of measurement is estimated indirectly, by analyzing the accuracy of measuring instruments, measurement conditions, and the obtained measurement data. In single measurements under reference condition of the instruments involved, the measurement error is determined by the limits of the permissible error of the instruments and is expressed in the form of limits of measurement error. In multiple measurements, the measurement inaccuracy is usually estimated using statistical methods, in which case the measurement inaccuracy is characterized using the concept of measurement uncertainty rather than the limits of error. The proposed definition of the term "error" is close to that given in [10] and also in [12].

The definition of *uncertainty* in VIM (clause 2.26) is provided with a note saying that uncertainty "may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability." This note creates ambiguity that is unacceptable in scientific terminology. Indeed, what is the uncertainty, a standard deviation or an interval? Giving two different meanings to one term must be avoided in a terminological dictionary.

9.3 Drawbacks of the "Guide to the Expression of Uncertainty in Measurement"

Another important document published by ISO is the "Guide to the Expression of Uncertainty in Measurement" (GUM) [2]. The goal of GUM was to unify the methods of measurement uncertainty estimation and its presentation. The uniformity of estimation and expression of inaccuracy of measurements is a necessary condition for the economic development of every country and for international economic cooperation. Thus, GUM was enthusiastically received by the metrological community.

However, a number of shortcomings among GUM recommendations have transpired subsequently. In [16], it was noted that "the evaluation methods in the GUM are applicable only to linear or linearized models and can yield unsatisfactory results in some cases." The same article reported that to address these issues, Addition 1 to GUM had been prepared and that furthermore, Working Group 1 JCGM decided in 2006 to prepare a new edition of GUM. Other critical comments regarding GUM can be found in [31]. Our own criticism appeared in [44] and, in more detail, in [42].

Still, the recently published VIM (which we discussed in the previous section) clearly reflects GUM's influence. For example, VIM repeatedly uses the notion of a set of true values of a measured quantity, which as we showed in Sect. 9.2 is misguided. In Note 3 to clause 2.11 it makes an attempt to justify a mistaken concept from GUM about the equivalency of the true value and a value of a quantity. Apparently, past criticisms of GUM were not sufficiently convincing, and we revisit its drawbacks here.

1. Scope of GUM

GUM begins with a statement that "The Guide establishes general rules for evaluating and expressing uncertainty in measurement that can be followed at various levels of accuracy and in many fields – from shop floor to fundamental research." Unfortunately, the rest of GUM's content does not support this intended scope since it is devoted exclusively to multiple measurements. Single measurements, although being the basic type of measurements in industry, trade, quality assurance, clinical medicine, and other fields, are not even mentioned. This limited scope is a significant limitation of GUM.

2. Philosophy of GUM

The foundational premise of GUM is that the concept of true value of a measurand is not needed because it is equal to the value of this measurand. This premise is formulated explicitly in "Guide Comment to Sect. B.2.3" (page 32 of GUM) and also in Annex D (Sect. D.3.5). However, this premise is in contradiction with VIM, as well as with fundamental conventions of physics and statistics. According to VIM, clause 1.19, the value of a quantity is a number and reference together expressing the magnitude of a quantity. In other words, it is the product of a number and the unit of measurement. This value is obtained as the result of a measurement. In contrast, the true value is a purely theoretical concept and cannot be found (see clause 2.11 of the VIM). Thus, the terms "true value" and "value of a quantity" cannot be considered the same and the latter cannot replace the former.

In statistics, the terms "parameter" (true value) and "estimate of the parameter" (the obtained value of the parameter) are strictly distinguished. In physics, the equations between physical quantities would be impossible without the concept of a true value; indeed, physical equations would always be only approximately correct for obtained values of the quantities. Finally, as we will see bellow, the GUM itself needed a distinction between the true value and the value of the measurand and was forced to introduce rather awkward new terminology in its place. These considerations bring a conclusion that during the new edition of GUM it should revert to traditional philosophy.

3. Terminology of the GUM

The elimination of the term "true value" was motivated by the desire to eliminate the term "error." Consequently, the GUM uses the term "uncertainty" in place of "error" throughout the document. The goal was to eliminate synonymia in using both terms throughout the document. This goal can be accomplished, however, without excluding the term "true value" and the corresponding concept; in fact, by defining the terms "error" and "uncertainty" precisely, we could distinguish the two clearly and at the same time not impoverish the metrological language by eliminating the term "error" but, to the contrary, enrich it by giving the two terms different meaning.

Metrology offers every prerequisite to achieve this. Indeed, the uncertainty of a measurement result is calculated usually from its components and with the help of statistical methods. In contrast, in the case of a single measurement using measurement instruments under reference conditions, the measurement inaccuracy is fully determined by the limits of error of the instrument, and statistical methods are not applicable.

Consequently, the term "uncertainty" may be used for probabilistic estimates of inaccuracy and the term "limits of error" when the inaccuracy estimates have no probabilistic interpretation. Moreover, according to VIM clause 2.26, the term "uncertainty" is associated with the result of measurement. Thus, it cannot replace the term "error" in other cases; for example, it cannot be used for components of uncertainty or to express the inaccuracy of a measuring instrument. We conclude that the total replacement of "error" with "uncertainty" is unjustified.

The GUM introduces two new terms "type A and type B evaluation of uncertainty," defining them as methods of evaluation of uncertainty (clause 2.3.2 and 2.3.3) but using them as components of uncertainty. Indeed, clause 5.1.2 describes how to combine uncertainties type A and type B; clearly, methods cannot be combined and they are treated there as components of uncertainty in this context. Such inconsistency should be avoided in a document aiming to introduce rigorous language for others to follow. In addition, these terms are not expressive. It would be much better to use the common term "random error" instead of "type A uncertainty" and the term "rated error" (if the term "systematic error" is undesirable).

Another inconsistency in the GUM is with the terms "standard uncertainty," "combined uncertainty," and "expanded uncertainty." The first two are defined as simply standard deviation and the combined standard deviation, respectively. But "expanded uncertainty" is presented as an interval. It is confusing to use the same term "uncertainty" as the basis for derived terms having drastically different meaning – a standard deviation in one case and an interval in the other.

In general, to calculate measurement uncertainty, the terms "standard deviation," "combined standard deviation," and "uncertainty" itself would be sufficient. The GUM introduced duplicate terms "standard uncertainty" and "combined standard uncertainty" as the terms that "are used sometimes for convenience" (clause 4.2.3). But it uses them exclusively throughout the rest of the document, creating an impression that this is the proper terminology to be used. These duplicate terms cause inconvenience in practice. For example, to follow this terminology, one has to always point out that standard uncertainty is equal to standard deviation, which is then computed using known statistical methods. As a typical example, Kacker and Jones [31] repeatedly use in their article passages the following: "According to the ISO Guide (Sect. 4.2), the type A standard uncertainty associated with z_A from classical statistics is $u(z_A) = s(z_A) = s(z)/\sqrt{m}$."

In other words, when saying "standard uncertainty," a methrologist must remember that in fact the term refers to "standard deviation." The same holds for the term "combined standard deviation."

Another terminological difficulty has to do with the concept of confidence interval. As it is known, it is the interval that, with given probability, contains the true value. Thus, it needs the concept of true value, which the GUM was trying to eliminate. In an attempt to resolve this logical gap, the GUM replaces the term "true value" with the expression "letter Y that represents the value attributed to the measurand" (clause 6.2.1 and Annex G) or "measurand Y." This proliferation of nondescriptive terms makes the terminology nonintuitive, and it is unnecessary since descriptive terms exist.

4. Evaluation of the uncertainty in the GUM

The GUM contains the terms standard uncertainty, combined uncertainty, and expanded uncertainty. The first two are just different names for the standard deviation and combined standard deviation. They are computed using known formulas. But expanded uncertainty is an interval. In Chap. 6 of the GUM this interval is called *coverage interval*, which is defined as "an interval about the measurement result that encompasses a large fraction p of the probability distribution of values that could reasonably be attributed to the measurand" (clause 6.1.2). GUM further describes the calculation procedure for the coverage interval using two additional new terms, *coverage probability* and *coverage factor*. The former has the same meaning as "confidence probability" and the latter as "quantile of the distribution." However, how to find the coverage factor and therefore the coverage interval remains unspecified and is unknown. Changing the terminology obviously does not solve the problem of obtaining the expanded uncertainty (or confidence interval in the traditional terminology).

The root of the problem with computing the expanded uncertainty is that the GUM does not provide a method for combining systematic and random errors of a measurement result. Consequently, clause 6.3.3 recommends calculating the expanded uncertainty simply as the product of combined uncertainty and factor 2 or 3; the result is assigned, without any justification, probability 0.95 in the first case and 0.99 in the second.

Clauses G.3.1 and G.3.2 of Annex G offer a different method for calculating the expanded uncertainty. This method is based on the Student's distribution, which in this case is not applicable. Indeed, recall that estimate of combined variance is a sum of estimates of variance of random errors (uncertainty A according to GUM) and conditional constant errors (uncertainty B). Thus, the combined standard deviation represents the standard deviation of the sum of random and conditionally constant systematic errors. Student's distribution establishes the connection between the mean of a group of observations and the standard deviation of this mean. In the case in question, the mean is calculated using data having only random errors, while the standard deviation – the square root of the sum of the estimates of the variances of random and conditionally constant errors – reflects both random and systematic errors. Therefore, using Student's distribution in this case is incorrect.

Another mistake has to do with calculating the effective degree of freedom. Its essence is that the concept of "degree of freedom" is not applicable to a random quantity with fully known distribution function. For the model of systematic errors the GUM takes the uniform distribution with known limits, and this distribution cannot be assigned degree of freedom $\nu = 1$, or any other number.

We should note that there is a known method for computing the uncertainty of a measurement result with given confidence probability, which accounts for both systematic and random errors of the result. This method is described in [44, 46] and discussed in detail in the present book.

The forgoing discussion shows that the upcoming new edition of the GUM must extend beyond revising its philosophy and terminology and revise its recommendations for data processing as well. Such revision is possible on the basis of existing methods and traditional philosophical foundation.

The revision of the GUM should utilize the method of reduction for dependent indirect measurements. In fact, the GUM already mentions the method of reduction as a second approach (see the note on page 10 in Sect. 4.1.4), but does not discuss its advantages over the used traditional method. These advantages were pointed out in this book, and the main ones being that this method allows one to construct the confidence interval for dependent indirect measurements and that it eliminates the need for the correlation coefficient. These benefits of the method of reduction are hard to overstate.

Further, we would like to point out again that the revision of the GUM must also include methods of estimating the inaccuracy of single measurements. These methods also exist already and are discussed in this book.

The above problems with GUM's recommendations regarding the estimation of the uncertainty of a measurement result have been recognized by JCGM, and Supplement 1 to the GUM is devoted to rectifying these issues [13]. Supplement 1 addresses them through the use of the Monte Carlo method. However, as we discussed earlier, there exist much simpler approaches. Note that being able to solve these problems without the Monte Carlo method would not obviate the need for Supplement 1 in the form of a separate recommendation devoted expressly to the Monte Carlo method, which can have its own significance in metrology (see Sect. 5.10).

Appendix

Table A.1 Values of the normalized Gaussian function $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-y^2/2} dy$

z	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34850	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42786	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861

Note: The values of $\Phi(z)$ for $z = 3.0$–4.5 are as follows:

3.0	0.49865	3.4	0.49966	3.8	0.49993
3.1	0.49903	3.5	0.49977	3.9	0.49995
3.2	0.49931	3.6	0.49984	4.0	0.499968
3.3	0.49952	3.7	0.49989	4.5	0.499997

Table A.2 Quantiles of Student's distribution

Degree of freedom ν	Significance level $q = (1-\alpha) \times 100(\%)$		
	10	5	1
1	6.31	12.71	63.66
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.90	2.36	3.50
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17
12	1.78	2.18	3.06
14	1.76	2.14	2.98
16	1.75	2.12	2.92
18	1.73	2.10	2.88
20	1.72	2.09	2.84
22	1.72	2.07	2.82
24	1.71	2.06	2.80
26	1.71	2.06	2.78
28	1.70	2.05	2.76
30	1.70	2.04	2.75
∞	1.64	1.96	2.58

Table A.3 Critical values of the distribution of $T_n = (x_n - \bar{x})/S$ or $T_1 = (\bar{x} - x_1)/S$ (with unilateral check)

Number of observations, n	Upper 0.5% significance level	Upper 1% significance level	Upper 5% significance level
3	1.155	1.155	1.153
4	1.496	1.492	1.463
5	1.764	1.749	1.672
6	1.973	1.944	1.822
7	2.139	2.097	1.938
8	2.274	2.221	2.032
9	2.387	2.323	2.110
10	2.482	2.410	2.176
11	2.564	2.485	2.234
12	2.636	2.550	2.285
13	2.699	2.607	2.331
14	2.755	2.659	2.371
15	2.806	2.705	2.409
16	2.852	2.747	2.443
17	2.894	2.785	2.475
18	2.932	2.821	2.504
19	2.968	2.854	2.532
20	3.001	2.884	2.557
21	3.031	2.912	2.580
22	3.060	2.939	2.603
23	3.087	2.963	2.624
24	3.112	2.987	2.644
25	3.135	3.009	2.663
26	3.157	3.029	2.681
27	3.178	3.049	2.698
28	3.199	3.068	2.714
29	3.218	3.085	2.730
30	3.236	3.103	2.745

Table A.4 Percentile points of the χ^2 distribution $P\left\{\chi^2 > \chi_q^2\right\}$

Degree of freedom ν	Significance level q (%)									
	99	95	90	80	70	30	20	10	5	1
1	0.00016	0.00393	0.0158	0.0642	0.148	1.074	1.642	2.706	3.841	6.635
2	0.0201	0.103	0.211	0.446	0.713	2.408	3.219	4.605	5.991	9.210
3	0.115	0.352	0.584	1.005	1.424	3.665	4.642	6.251	7.815	11.345
4	0.297	0.711	1.064	1.649	2.195	4.878	5.989	7.779	9.488	13.277
5	0.554	1.145	1.610	2.343	3.000	6.064	7.289	9.236	11.070	15.086
6	0.872	1.635	2.204	3.070	3.828	7.231	8.558	10.645	12.592	16.812
7	1.239	2.167	2.833	3.822	4.671	8.383	9.803	12.017	14.067	18.475
8	1.646	2.733	3.490	4.594	5.527	9.524	11.030	13.362	15.507	20.090
9	2.088	3.325	4.168	5.380	6.393	10.656	12.242	14.684	16.919	21.666
10	2.558	3.940	4.865	6.179	7.267	11.781	13.442	15.987	18.307	23.209
11	3.053	4.575	5.578	6.989	8.148	12.899	14.631	17.275	19.675	24.725
12	3.571	5.226	6.304	7.807	9.034	14.011	15.812	18.549	21.026	26.217
13	4.107	5.892	7.042	8.634	9.926	15.119	16.985	19.812	22.362	27.688
14	4.660	6.571	7.790	9.467	10.821	16.222	18.151	21.064	23.685	29.141
15	5.229	7.261	8.547	10.307	11.721	17.322	19.311	22.307	24.996	30.578
16	5.812	7.962	9.312	11.152	12.624	18.418	20.465	23.542	26.296	32.000
17	6.408	8.672	10.085	12.002	13.531	19.511	21.615	24.769	27.587	33.409
18	7.015	9.390	10.865	12.857	14.440	20.601	22.760	25.989	28.869	34.805
19	7.633	10.117	11.651	13.716	15.352	21.689	23.900	27.204	30.144	36.191
20	8.260	10.851	12.443	14.578	16.266	22.775	25.038	28.412	31.410	37.566
21	8.897	11.591	13.240	15.445	17.182	23.858	26.171	29.615	32.671	38.932
22	9.542	12.338	14.041	16.314	18.101	24.939	27.301	30.813	33.924	40.289
23	10.196	13.091	14.848	17.187	19.021	26.018	28.429	32.007	35.172	41.638
24	10.856	13.848	15.659	18.062	19.943	27.096	29.553	33.196	36.415	42.980
25	11.524	14.611	16.473	18.940	20.867	28.172	30.675	34.382	37.652	44.314
26	12.198	15.379	17.292	19.820	21.792	29.246	31.795	35.563	38.885	45.642
27	12.879	16.151	18.114	20.703	22.719	30.319	32.912	36.741	40.113	46.963
28	13.565	16.928	18.939	21.588	23.647	31.391	34.027	37.916	41.337	48.278
29	14.256	17.708	19.768	22.475	24.577	32.461	35.139	39.087	42.557	49.588
30	14.953	18.493	20.599	23.364	25.508	33.530	36.250	40.256	43.773	50.892

Table A.5 Values of the upper 1% of points of the distribution $F_{0.01} = S_1^2/S_2^2$

Degree of freedom

ν_2	ν_1 2	3	4	5	6	8	12	16	24	50	∞
2	99.00	99.17	99.25	99.30	99.33	99.36	99.42	99.44	99.46	99.48	99.50
3	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.83	26.60	26.35	26.12
4	18.00	16.69	15.98	15.52	15.21	14.80	14.37	14.15	13.93	13.69	13.46
5	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.68	9.47	9.24	9.02
6	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.52	7.31	7.09	6.88
7	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.27	6.07	5.85	5.65
8	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.48	5.28	5.06	4.86
9	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.92	4.73	4.51	4.31
10	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.52	4.33	4.12	3.91
11	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.21	4.02	3.80	3.60
12	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.98	3.78	3.56	3.36
13	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.78	3.59	3.37	3.16
14	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.62	3.43	3.21	3.00
15	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.48	3.29	3.07	2.87
16	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.37	3.18	2.96	2.75
17	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.27	3.08	2.86	2.65
18	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.20	3.00	2.79	2.57
19	5.93	5.01	4.50	4.17	3.94	3.63	3.30	3.12	2.92	2.70	2.49
20	5.85	4.94	4.43	4.10	3.87	3.56	3.23	3.05	2.86	2.63	2.42
21	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.99	2.80	2.58	2.36
22	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.94	2.75	2.53	2.31
23	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.89	2.70	2.48	2.26
24	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.85	2.66	2.44	2.21
25	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.81	2.62	2.40	2.17
26	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.78	2.58	2.36	2.13
27	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.74	2.55	2.33	2.10
28	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.71	2.52	2.30	2.06
29	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.68	2.49	2.27	2.03
30	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.66	2.47	2.24	2.01
35	5.27	4.40	3.91	3.59	3.37	3.07	2.74	2.56	2.37	2.13	1.90
40	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.48	2.29	2.05	1.80
45	5.11	4.25	3.77	3.45	3.23	2.94	2.61	2.43	2.23	1.99	1.75
50	5.06	4.20	3.72	3.41	3.19	2.89	2.56	2.38	2.18	1.94	1.68
60	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.32	2.12	1.87	1.60
70	4.92	4.07	3.60	3.29	3.07	2.78	2.45	2.28	2.07	1.82	1.53
80	4.88	4.04	3.56	3.26	3.04	2.74	2.42	2.24	2.03	1.78	1.49
90	4.85	4.01	3.53	3.23	3.01	2.72	2.39	2.21	2.00	1.75	1.45
100	4.82	3.98	3.51	3.21	2.99	2.69	2.37	2.19	1.98	1.73	1.43
125	4.78	3.94	3.47	3.17	2.95	2.66	2.33	2.15	1.94	1.69	1.37
∞	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.99	1.79	1.52	1.00

Table A.6 Values of the upper 5% of points of the distribution $F_{0.05} = S_1^2/S_2^2$

Degree of freedom											
	ν_1										
ν_2	2	3	4	5	6	8	12	16	24	50	∞
2	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.43	19.45	19.47	19.50
3	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.69	8.64	8.58	8.53
4	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.84	5.77	5.70	5.63
5	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.60	4.53	4.44	4.36
6	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.92	3.84	3.75	3.67
7	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.49	3.41	3.32	3.23
8	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.20	3.12	3.03	2.93
9	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.98	2.90	2.80	2.71
10	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.82	2.74	2.64	2.54
11	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.70	2.61	2.50	2.40
12	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.60	2.50	2.40	2.30
13	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.51	2.42	2.32	2.21
14	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.44	2.35	2.24	2.13
15	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.39	2.29	2.18	2.07
16	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.33	2.24	2.13	2.01
17	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.29	2.19	2.08	1.96
18	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.25	2.15	2.04	1.92
19	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.21	2.11	2.00	1.88
20	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.18	2.08	1.96	1.64
21	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.15	2.05	1.93	1.81
22	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.13	2.03	1.91	1.78
23	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.11	2.00	1.88	1.76
24	3.40	3.01	2.78	2.62	2.51	2.36	2.18	2.09	1.98	1.86	1.73
25	3.38	2.99	2.76	2.60	2.49	2.34	2.16	2.07	1.96	1.84	1.71
26	3.37	2.98	2.74	2.59	2.47	2.32	2.15	2.05	1.95	1.82	1.69
27	3.35	2.96	2.73	2.57	2.46	2.30	2.13	2.03	1.93	1.80	1.67
28	3.34	2.95	2.71	2.56	2.44	2.29	2.12	2.02	1.91	1.78	1.65
29	3.33	2.93	2.70	2.54	2.43	2.28	2.10	2.00	1.90	1.77	1.64
30	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.99	1.89	1.76	1.62
35	3.26	2.87	2.64	2.48	2.37	2.22	2.04	1.94	1.83	1.70	1.57
40	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.90	1.79	1.66	1.51
45	3.21	2.81	2.58	2.42	2.31	2.15	1.97	1.87	1.76	1.63	1.48
50	3.18	2.79	2.56	2.40	2.29	2.13	1.95	1.85	1.74	1.60	1.44
60	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.81	1.70	1.56	1.39
70	3.13	2.74	2.50	2.35	2.23	2.07	1.89	1.79	1.67	1.53	1.35
80	3.11	2.72	2.49	2.33	2.21	2.06	1.88	1.77	1.65	1.51	1.32
90	3.10	2.71	2.47	2.32	2.20	2.04	1.86	1.76	1.64	1.49	1.30
100	3.09	2.70	2.46	2.30	2.19	2.03	1.85	1.75	1.63	1.48	1.28
125	3.07	2.68	2.44	2.29	2.17	2.01	1.83	1.72	1.60	1.45	1.25
∞	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.64	1.52	1.35	1.00

Glossary

Absolutely constant error An elementary error of a measurement that remains the same in repeated measurements performed under the same conditions. The value of this error is unknown, but its limits can be estimated.

Examples: (1) An error of indirect measurement caused by using imprecise equation between the measurand and measurement arguments. (2) An error in voltage measurement that uses a moving-coil voltmeter when the resistance of the voltage source is unknown.

Accuracy class A class of measuring devices that meets stated metrological requirements. Accuracy classes are intended to optimize the number of different accuracy levels of measuring devices and to keep their errors within specified limits.

Accuracy of measurement Closeness of the result of measurement to the true value of the measurand.

Accuracy of measuring instrument The ability of a measuring instrument to perform measurements with results that is close to the true values of the measurands.

Additional error of measuring instrument The difference between the error of a measuring instrument when the value of one influence quantity exceeds its reference value and the error of that instrument under reference condition.

Argument influence coefficient Partial derivative of the function at the right-hand side of the measurement equation of an indirect measurement with respect to one argument.

Notes: (1) Argument influence coefficient is calculated by substituting the arguments in the resulting derivative function with their estimates. (2) Argument influence coefficients are expressed in absolute or relative form.

Calibration Operation that, under specified conditions, establishes the relationship between values indicated by a measuring instrument and corresponding values obtained from a measurement standard.

Notes: (1) Results of calibration may be presented by a table, calibration curve or by a table of additive or multiplicative corrections of the instrument or measure indications. (2) The ratio of permissible error limits for measuring instrument or measure being calibrated and uncertainty of measurement standard are stated in national or international recommendations or standards or it is adopted by calibration laboratories and may be different in different fields of measurement.

Conditionally constant error An unknown elementary error of a measurement that lies inside an interval defined by the known limits of permissible error of the measuring instrument involved.

Note: The limits of permissible error are the same for all measuring instruments of particular type and therefore those instruments are interchangeable in that sense.

Dead band An interval through which a stimulus signal at the input of measuring instrument may be changed without response in instrument indication.

Direct measurement A measurement in which the value of the measurand is read from the indication of the measuring instrument; the latter can be multiplied by some factor or adjusted by applying certain corrections.

Dynamic measurement A measurement in which the measuring instrument is employed in dynamic regime.

Drift A slow change in output indication of a measuring instrument that is independent of a stimulus.

Note: The drift is usually checked at the zero point of a measuring instrument indication and is eliminated by adjusting the instrument indication to the zero point before measurement.

Elementary error of a measurement A component of error or uncertainty of a measurement associated with a single source of inaccuracy of the measurement.

Error of a measurement A deviation of the result of a measurement from the true value of the measurand.

Note: Error of measurement may be expressed in absolute or relative form.

Fiducial error A ratio of the permissible limits of the absolute error of the measuring instrument to some standardized value – *fiducial value*. Fiducial error is expressed as percentage and makes it possible to compare the accuracy of measuring instruments that have different measurement ranges and different limits of permissible error when the latter are expressed in absolute form.

Fiducial value Quantity value specified for a particular type of measuring instruments. Fiducial value may be, for example, the span or the upper limit of the nominal range of the measuring instrument.

Inaccuracy of a measurement A quantitative characteristic of the degree of deviation of a measurement result from the true value of the measurand.

Note: Inaccuracy of a measurement may be expressed as limits of measurement error or as measurement uncertainty.

Indirect measurement A measurement in which the estimate of the measurand is calculated using measurements of other quantities related to the measurand by known function.

Influence coefficient A factor that after multiplying by a value of deviation of a specific influence quantity from its reference condition limits gives the additional error.

Influence function A metrological characteristic of the measuring instrument expressing the relationship between errors of that instrument and values of an influence quantity.

Intrinsic error The error of a measuring instrument determined under reference conditions.

Limits of measurement error Limits of the deviation of the measurement result from the true value of the measurand.

Limits of permissible error of a measuring instrument Maximum value of an error that is permitted by specification for a given measuring instrument.

Material measure A measuring instrument that reproduces a particular kind of quantity with known value and accuracy.
Note: The indication of a material measure is its assigned quantity value.

Measurand A particular quantity whose value must be obtained by measurement.

Measurement A set of experimental operations, involving at least one measuring instrument, performed for the purpose of obtaining the value of a quantity.

Measurement chain A set of several measuring instruments connected temporary in a chain to perform a measurement.

Measurement standard A measuring instrument intended to materialize and/or conserve a unit of a quantity in order to transfer its value to all other measuring instruments.
Note: There are primary measurement standard, secondary standards, standards with specified functions and at the end of this chain – working standards.

Measurement vector A set of matched measurements of all arguments defining an indirect measurement.

Measuring instrument A technical product that is created for the purpose to be used in a measurement and which has known metrological characteristics.

Metrological characteristic A characteristic of a measuring instrument that allows one to judge the suitability of the instrument for measurement in a given range, or that is necessary for the estimation of the inaccuracy of measurement results.

Metrology Science of measurements regardless of the field to which the quantity to be measured belongs and of the accuracy of measurements.

Primary measurement standard A measurement standard that has the highest accuracy in a country.

Note: The primary measurement standard usually is recognized by national authority as national standard and used for assigning the measurement unit to other measurement standards for the kind of quantity concerned.

Random error A component of the inaccuracy of a measurement that, in the repeated measurements of the same measurand under the same conditions, varies in an unpredictable way.

Rated conditions Operating conditions, determined for specified type of measuring instruments, that are wider than their reference operating conditions and nevertheless allow the estimation of the inaccuracy of a measurement performed by this type instrument under these conditions.

Note: Rated conditions are described as permissible excess value of influence quantities over those given as limits for reference conditions.

Reference conditions Operating conditions, determined for specified type of measuring instruments, under which the measurement performed by this type instrument is more accurate than under other conditions.

Repeatability of a measurement Agreement among several consecutive measurements for the same measurand performed, under the same operating conditions with the same measuring instruments, over a short period of time.

Reproducibility of a measurement Agreement among measurements for the same measurand performed in different locations, under different operating conditions, or over a long period of time.

Response time The time interval between the instant when a measuring instrument gets a stimulus and the instant when the response reaches and remains within specified limits of its final steady value.

Result of measurement The value of a measurand obtained by measurement.

Note: The measurement result is expressed as a product of a number and a proper unit.

Secondary measurement standard A measurement standard that obtains the magnitude of a unit from the primary measurement standard.

Span The absolute value of the difference between the two limits of the nominal range of a measuring instrument.

Example: Voltmeter with the nominal range from -15 to $+15$ V has the span of 30 V.

Systematic error A component of the inaccuracy of a measurement that, in the repeated measurements of the same measurand under the same conditions, remains constant or varies in a predictable way.

True value of the measurand The value of a quantity that being known would ideally reflect the property of an object with respect to the purpose of the measurement.

Note: True value can never be found.

Uncertainty of measurement An interval within which a true value of a measurand lies with given confidence probability.

Notes: (1) Uncertainty is expressed by its limits, which are listed as offsets from the result of the measurement. (2) Uncertainty may be presented either in absolute or relative form.

Verification A kind of calibration that reveals whether the error of a measuring instrument lies within their permissible limits.

Working standard A measurement standard that is used to calibrate measuring instruments.

References

Standards and Recommendations

1. International vocabulary of metrology – Basic and general concepts and associated terms (VIM), ISO/IEC Guide 99 (2007).
2. Guide to the Expression of Uncertainty in Measurement, ISO (1995).
3. Standard Practice for Use of the International System of units (SI), American Society for Testing and Materials E 380–91 (1991).
4. Standard Practice for Dealing with Outlying Observations, American Society for Testing and Materials, E 178–02 (Reapproved 2002).
5. National Measurement Accreditation Service, B 3003, English National Physical Laboratory, Teddington (1986).
6. Measurement Uncertainty, American National Standard ANSI/ASMEPTC19.1 (1985).
7. Standard Practice for the Evaluation of Single-Pan Mechanical Balances, American Society for Testing and Materials, E 319–85 (1985).
8. Publication 51, PT 1–84, Direct Acting Indicating Analogue Electrical Measuring Instruments and Their Accessories, 4th ed., Bureau Central de la Comission Electrotechnique Internationale, Geneva (1984).
9. Recommandation Internationale 34, Classes de precision des Instruments de Mesurage, Organisation International de Metrologie Legale, Paris (1974).
10. Process Instrumentation Terminology, American National Standard ANSI/ISA S51.1 (1979).
11. Direct Measurements with Multiple Observations, Methods of Processing the Results of Observations: Basic principles, State Standard of the Soviet Union, GOST 8.207–76, Moscow (1976).
12. Metrology, Terms and Definitions, State Standard of the Soviet Union, GOST16263–70, Moscow (1970).
13. Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement". Propagation of distributions using a Monte Carlo method. Final draft. Joint Committee for Guides in Metrology (JCGM). September 2006.

Books and Articles

14. R. B. Abernethy et al., *Measurement Uncertainty Handbook* (Instrument Society of America 1980).
15. V. Ya. Alekseev, F.M. Karavaev, Zn.F. Kudryshova, S.G. Rabinovich. "Estimate of the measured quantity from results of measurements with different systematic errors," *Metrologia* (Supplement to the journal *Izmeritel'naya Technika*[1]) 1, 18–24 (1978).

[1] The journal *Izmeritel'naya Technika* was published for many years in the United States in English under the title *Measurement Techniques*.

16. W. Bich, M.G. Cox, P.M. Harris. "Evolution of the "Guide to the expression of uncertainty in measurement"," *Metrologia* 43, 161–166 (2006).
17. H. Bäckström, "Uber die dezimalgleichung beim ablesen von scalen," *Zeitschrift furInstrumentenkunde* (1930–32).
18. J.L. Bucher (Editor), *Metrology Handbook* (ASQ Quality Press, Wilwaukee, WI 2004).
19. H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
20. C. Croarkin, *Measurement Assurance Programs,* Part II: *Development and Implementation*, NBS Special Publication 676–11 (U.S. GPO, Washington, DC, 1985).
21. G. D'Agostini, *Bayesian Reasoning in Data Analysis. A Critical Introduction*, (World Scientific Publishing Co, River Edge, NJ, 2003).
22. R.H. Dieck, *Measurement Uncertainty. Methods and Applications*, (ISA, Research Triangle Park, NC, 1992).
23. B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap*, (Chapman & Hall/CRC Press, 1998).
24. B.D. Ellis, *Basic Concepts of Measurement*, (Cambridge University Press, Cambridge, 1966).
25. C. Eisenhart, "Realistic evaluation of the precision and accuracy of instrument calibration systems," *Journal of Research of the National Bureau of Standards* 67C, 161–187 (1963).
26. Fluke, *Calibration: Philosophy in Practice,* Fluke Corporation, 2nd ed. Everett, P.O. Box 9090, WA 98206, (1994).
27. V.A. Granovskii, *Dynamic Measurements, Principles of Metrological Support* [in Russian] (Energoatomizdat, Leningrad, 1984).
28. V.A. Granovskii, T.N. Siraya, *Methods of Processing of the Measurements of Experimental Data* [in Russian] (Energoatomizdat, Leningrad division, Leningrad, 1990).
29. R.V. Hogg, "Some observations on robust estimation," *Journal of the American Statistical Association* 62(320), 1179–1186 (1967).
30. R. Kaarls, "Metrology, essential to trade, industry and society," *Accreditation and Quality Assurance* 12(8), 435–437 (2007).
31. R. Kacker, A. Jones, "On use of Bayesian statistics to make the "Guide to the expression of uncertainty in measurement" consistent," *Metrologia* 40, 235–248 (2003).
32. A.M. Kagan, Yu.V. Linnik, "On some problems of mathematical statistics in metrology," in *Fifty Years of Metric Reform in the USSR,* Proceedings of the Metrological Institute of the USSR [in Russian], D.I. Mendeleev All-Union Scientific-Research Institutes of Metrology, No. 123 (183) (Leningrad, 1970), pp. 39–47.
33. F.M. Karavaev, *Measurements of the Activity of Nuclides* [in Russian] (Izd. standartov, Moscow, 1972).
34. Zn.F. Kudryshova, S.G. Rabinovich, "Methods of experimental data processing in indirect measurements," in *Methods of Experimental Data Processing in Measurements*, Proceedings of the Metrological Institutes of the USSR [in Russian], D.I. Mendeleev All-Union Scientific Institute of Metrology, No. 172 (232), (Energia, Leningrad, 1975).
35. A.M. Law, W.D. Kelton, Simulation modeling and analysis. McGraw-Hill, Inc, New York, (1982).
36. I. Lira, *Evaluating the measurement uncertainty. Fundaments and practical guidance.* With an introductory chapter by T.J. Quinn, Director of the BIPM. Institute of Physics Publishing, (2002).
37. M.F. Malikov, *Foundations of Metrology* [in Russian] (Committee on Measures and Measuring Devices at the Council of Ministers of the USSR, Moscow, 1949).
38. B.S. Massey, *Measures in Science and Engineering, Their Expression, Relation and Interpretation* (Wiley, New York, 1986).
39. A.I. Mechanikov, "Methods for immediate processing of the results of observations with the help of ordering statistics," *Metrologia* (Supplement to *Izmeritel'naya Technika*) 4, 47–58 (1972).
40. P.J. Mohr, B.N. Taylor, "CODATA recommended values of fundamental physical constants: 2002" *Review of Modern Physics* 77(1) 1–107, (2005).

41. J.G. Nee, *Fundaments of tool design. 4-th ed*. Society of manufacturing engineers, (1998).
42. S. Rabinovich, "Towards a new edition of the "Guide to the expression of uncertainty in measurement"," *Accreditation and Quality Assurance* 12(11) 603–608, (2007).
43. S. Rabinovich, "Accuracy of single measurements," *Accreditation and Quality Assurance* 12(8) 419–424, (2007).
44. S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, 3rd ed (Springer, New York, 2005).
45. S.G. Rabinovich, "An efficient calculation for indirect measurements and a new approach to the theory of indirect measurements," *Proceedings of the Measurement Science Conference*, Anaheim, California, January 25 and 26, 1996.
46. S.G. Rabinovich, *Measurement Errors* [in Russian] (Energia, Leningrad, 1978).
47. S.G. Rabinovich and T.L. Yakovleva, "Analysis of the temporal stability of the distributions of the errors of measuring instruments," *Metrologia* (Supplement to *Izmeritel'naya Technika)* 7, 8–15 (1977).
48. S.G. Rabinovich, "Method for calculating the measurement errors," *Metrologia* (Supplement to *Izmeritel'naya Technika*), No. 1, 3–12 (1970).
49. L. Sachs, *Applied Statistics: A Handbook of Techniques* (Springer-Verlag, New York, 1984).
50. K.P. Shirokov, *Basic Concepts of Metrology, Proceedings of the Metrological Institutes of the USSR* [in Russian], D.I. Mendeleev All-Union Scientific-Research Institute of Metrology No. 130 (190) (Izd. standartov, M.-L., 1972).
51. K.P. Shirokov, V.O. Arutyunov, E.M. Aristov, V.A. Granovskii, W.S. Pellinez, S.G. Rabinovich, D.F. Tartakovskii, "Basic concepts of the theory of dynamic measurements," *Proceedings of the 8th IMEKO Congress* (Akademiai Kiado, Budapest, 1980).
52. V.M. Sviridenko, "Logical-gnoseological aspect of the problem of the accuracy of measurements," *Izmeritel'naya Technika*1 5, 6–8 (1971).
53. B.L. van der Waerden, *Mathematical Statistics* (Springer-Verlag, New York, 1969).
54. T.L. Yakovleva, "About statistical homogeneity of the samples of measuring instrument errors," *Metrologiya* (Supplement to *Izmeritel'naya Technika*1) 2, 19–23 (1979).

Index

A

A posteriori estimation, 21, 22, 122, 215
A priori estimation, 21, 22, 107, 122, 209–211
Absolute error, 2, 3, 34, 37, 38, 40, 41, 45, 47, 120, 122, 162
Absolutely constant errors, 107, 109, 124, 126, 130, 146, 163, 165, 202, 204, 211
Accuracy, 17, 27, 28, 41, 56, 92, 115, 142, 144, 154, 156, 159, 162, 169, 171, 175, 177–179, 185, 187, 191, 192, 197, 207, 222, 225, 226, 229, 233, 240, 242–245
  analog instruments, 32
  basic concepts, 1–3
  Bayes' theorem, 99–100
  calibration, 53, 59–63
  characteristics, 20–23
  classes, 46–48
  cost considerations, 56
  definition, 9
  estimation, 101, 103–105, 239
  fiducial errors, 37
  inaccuracy, 25–26
  indirect measurements, 166–167
  measuring instruments, 15
  metrological characteristics, 33–36
  multiple measurements, 128–135
  single measurements, 119–127, 167–168
  stability, 45
Accuracy class index, 47
Accuracy classes of measuring instruments, 46, 47
Additional error, 19, 37, 39, 43–46, 110, 111, 119, 122, 123, 127, 164, 165, 167, 209, 211, 217, 235
Alternating current parameters, 15, 17
Arguments, 18, 20, 25, 27, 84, 99, 103, 109, 112, 113, 141, 142, 144–170, 191, 222, 223, 225–233, 236
Arithmetic mean, 75, 76, 85, 87, 91–93, 98, 119, 128–130, 132, 182, 195–197, 199–201

B

Bartlett's test, 90
Bias, 79, 150
Borda's weighing method, 117

C

Calibration
  curve, 188, 192
  schemes, 58–65
  verification, 53–59
Central limit theorem, 73, 114, 129
Chain of measurement instruments, 19, 60, 62, 167–168, 216
Chebyshev inequalities, 83, 84
Combined average, 194, 195, 204
Combined measurements, 18, 19, 174, 184, 200, 206, 207
Combined standard deviation, 130, 134, 152, 159, 165, 202, 203, 246, 247
Combined uncertainty, 165, 228, 232, 246, 247
Combining results of measurements, 170, 193–208
Conditional equations, 174–188, 190, 192
Conditionally constant errors, 107–110, 124, 126, 130–132, 134, 157–159, 163–165, 201, 211, 231, 232, 241, 247
Confidence interval, 2, 82–87, 97–99, 112, 114, 125, 129–134, 148, 149, 151, 152, 157, 158, 171, 177, 184, 191, 204–206, 231, 240, 247, 248
Confidence probability, 27, 80, 112–114, 116, 125, 130, 133, 134, 136, 155, 165, 184, 192, 201, 203, 205, 221, 235, 248
  component errors and, 114

confidence interval, 82–86, 98, 151–153
errors, 231, 232
GUM, 247
reference conditions, 167
single measurement, 131
Student's distribution, 149, 159, 226, 227, 238
systematic error, 158
Constraint equation, 178, 182
Constraints, 178, 182
Conventional notations, 41, 42, 47
Correlation coefficient, 142–144, 147, 153, 154, 224, 240, 248
Covariance, 143

D

Dead band, 35, 43, 214
Degree of freedom, 86–88, 130, 137, 148, 149, 152, 154, 157–160, 178, 221, 224–229, 231, 232, 238, 247
Derived physical quantities, 12, 186, 245
Digital instruments, 23, 32, 33, 41, 108
Direct measurement, 100, 128, 168, 184, 209, 216
Distribution, 100, 106, 149, 204, 206, 214, 238, 241
normal, 125, 128–138
parameters, 76–80
probabilistic analysis, 55, 56, 110, 148, 170
random errors, 83
testing hypotheses, 87–89
uniform, 111–117, 164, 167
Distribution function, 17, 27, 66–69, 71–76, 168, 169
central moments, 73–75
composition of, 111–117, 132, 133, 169, 229
excess, 75, 96, 97
initial moments, 74
mathematical expectation of random quantity, 16, 48, 72, 74, 75, 77–78, 93, 114, 129, 150, 169, 202
normal distribution, 67, 72, 73, 75–82, 84, 85, 87, 89, 93, 94, 98, 100, 110, 112, 114, 115, 125, 129, 132–135, 137, 151, 241
normalized Gaussian function, 73, 84, 91
probability density, 55, 71–77, 99, 111, 112, 242
probability distribution, 55, 56, 68, 72, 74, 75, 92, 97, 110, 112, 133, 148, 149, 170, 247
skewness, 75, 82
stability of, 67, 170
standard deviation of random quantity, 35, 48, 74, 79, 80, 83, 85, 86, 108, 110, 111, 114, 129–132, 148, 158, 165, 202, 203, 205, 206, 232, 235–237, 246, 247
uniform, 72, 73, 75, 76, 110–117, 129, 132–135, 137, 164, 167, 204, 207, 247
variance of random quantity, 16, 72, 74–76, 79, 87, 91, 92, 129, 130, 132, 143, 144, 150, 169, 202, 205
Dynamic characteristics
amplitude-frequency response, 50
amplitude-and phase-frequency response, 49
complete dynamic characteristics, 49, 52
differential equation, 20, 49–52
dynamic error, 24, 25, 45, 52, 105, 123
frequency response, 50, 52
frequency response of phase, 49
partial dynamic characteristic, 49, 52
response time, 49, 52, 102, 103
transfer function, 20, 34, 49, 50
transient response, 49, 51, 52

E

Effective degree of freedom, 137, 148, 149, 200, 201, 203, 228, 229, 232, 238, 247
Elementary errors, 23, 25, 104–111, 116, 119, 122–124, 127, 136, 164, 166, 167, 205, 206, 213–215
Elimination of systematic errors
method of symmetric observations, 119
progressing systematic errors, 106, 118
Error propagation formula, 148
Errors of measurement
absolute error, 2, 3, 34, 37, 38, 40, 41, 45, 47, 120, 122, 162
absolutely constant, 107, 109, 126, 130, 136, 146, 162, 163, 165, 201, 202, 204, 211
blunders, 24, 80
conditionally constant, 107–111, 124–126, 130–132, 134, 135, 146, 157–159, 163–165, 201, 202, 210, 211, 231, 232, 234–236, 241, 247
confidence, 132, 133, 231
confidence limit of random, 131, 137, 158, 206
confidence limits of, 85, 131, 133, 134, 137, 151, 158, 164, 190, 204–206, 208, 227, 231

constant systematic, 106, 107, 118, 132, 135, 137, 162, 201, 202, 204, 235, 236, 247
dynamic, 24, 25, 45, 52, 105, 123
elementary, 23, 25, 104–111, 116, 119, 122–124, 127, 136, 164, 166, 167, 205, 206, 213–215
gross, 24
individual estimation of, 21, 214–215
instrumental, 23, 47, 68, 105
instrumental systematic, 105
limit of permissible measurement error, 13, 14
methodological, 23, 105, 162, 201
outlying, 24, 80
periodic, 106, 109
personal, 23
personal systematic, 105
progressing, 106, 118
quasirandom, 108–111
random, 3, 23, 24, 27, 35, 39, 71, 83, 89, 102, 104, 105, 108, 110, 117, 124, 130–132, 136–139, 146–149, 157–159, 162, 165, 199–204, 206, 214, 217, 226, 230, 235, 236, 239–241, 246–248
regularly varying systematic errors, 106, 107
relative, 2, 3, 37, 39–42, 46, 119, 121, 122, 124, 133, 156, 162, 163, 198, 199, 210, 213, 219, 222, 243
static, 24
systematic, 3, 23, 24, 27, 39, 43, 92, 104–109, 117–119, 130, 132, 135–139, 146, 149, 150, 154, 157, 158, 161–164, 173, 174, 199–204, 220, 226, 231, 235, 236, 238–241, 246–248
universal estimation of errors, 21, 120, 212–213, 219
Errors of measuring instruments
additional errors, 19, 37, 39, 43–46, 110, 111, 119, 122, 123, 125, 127, 164, 165, 167, 209, 211, 217, 235
additional errors in dynamic regime, 45, 123
dynamic error, 24, 25, 45, 52, 105, 123
error of measures, 213, 236, 244
fiducial errors, 37, 38, 40, 41, 47, 119–121, 124, 239
intrinsic error, 23, 37, 43–44, 47, 108, 110, 119–124, 127, 130, 166, 207, 209, 210
permissible absolute error, 40, 41
permissible additional error, 43
permissible error, 13, 37, 40–42, 47, 54–58, 66, 206, 217, 218, 244
permissible relative error, 41–42
variation of indications, 33, 43
Estimates
consistent, 75–79, 128
efficient, 75, 76, 84, 128, 129, 194
unbiased, 75, 77–79, 90, 128, 153
Excess, 75, 96, 97
Expansion, 10, 145, 149, 185, 228

F
False rejection, 54, 55, 88, 89
False retention, 54, 55, 88
Fiducial error, 37, 38, 40, 41, 47, 119–121, 124, 239
Fiducial value, 37, 40, 41, 119, 120
Fisher's distribution, 92
Fisher's test, 90–92, 201

G
Gauss's method of weighing, 118
Gauss's notation, 176
General Conference on Weights and Measures (CGPM), 4

H
Histogram, 87, 88, 169, 170
Homogeneity, 66, 89–95

I
Inaccuracy of measurements
individual estimation of, 21, 214–215
universal estimation of, 21, 22, 120, 212–213, 219
Indirect measurement
dependent, 4, 32, 233
independent, 4, 32
linear, 148, 154
measurement, 18–20, 27, 100, 111, 141–171, 191, 204, 216, 221–226, 240, 248
measurement vector, 144, 147, 152–154, 222, 225
nonlinear, 142, 148–151, 154, 226
nonlinear independent, 149, 226
single, 14, 20, 24, 33, 39, 90, 101–103, 119–128, 130, 131, 137, 141, 161, 166–168, 181, 204–206, 216, 239, 240, 243–246, 248

Influence coefficients, 34, 43, 44, 146, 149, 157, 162–164, 166–168, 222, 224, 228, 231, 236
Influence function, 34, 44, 123
Influence quantities, 21, 23, 34, 37, 43–45, 49, 105, 108, 111, 122, 123, 125, 146, 164, 165, 167, 206
Input signal
  informative parameter of, 44
  noninformative parameters of, 44, 45, 105
Instrumental error, 23, 47, 68, 105
International Bureau of Weights and Measures (BIPM), 4, 11, 12
International Electrotechnical Commission (IEC), 37, 40, 41, 241
International Organization for Standardization (ISO), 136, 240, 241, 244, 246
International Organization of Legal Metrology (OIML), 37, 40–42, 46
International System of Units, 4, 6, 9
Intrinsic error, 23, 37, 43–44, 47, 108, 110, 119–124, 127, 130, 166, 207, 209, 210

K
Kolmogorov–Smirnov test, 87

L
Least-squares method, 173, 174, 187, 188, 190, 191, 241
Legendre's principle, 175
Limit of permissible instrument error, 56, 58, 207
Limiting random error, 35, 130, 131, 214
Limits of error, 2, 3, 22, 25, 27–29, 35, 39, 43, 45, 46, 56, 101, 109, 120, 121, 131, 157, 168, 204–206, 208, 213, 214, 218, 219, 222, 225, 234, 244, 246
Linear calibration curves, 188, 192
Linear conditional equations, 180, 181

M
Mathematical expectation of random quality, 74
Measurable quantity, 1, 2, 12–15, 17, 18, 21–24, 26, 32
Measurand, 19, 32, 44, 82, 95–99, 103, 125, 128, 129, 141, 145, 148–151, 153–155, 173, 229
  basic concepts, 1–3
  definition, 242, 244
  estimate of, 158, 160, 166, 170, 193, 204, 205, 221, 223, 228, 232, 236
  GUM, 245, 247
  indirect measurement, 191
  multiple measurements, 181
  properties, 24
  true value of, 13, 16, 18, 22, 26, 33, 35, 41, 156, 225
Measurements
  of activity of nuclides, 236
  arguments, 141, 153
  biasness, 79, 173
  combined, 18, 19, 174, 184
  dependent indirect, 142, 148, 149, 151–155, 161, 164, 169, 191, 221, 222, 225, 226, 240, 248
  direct, 18–20, 101–139, 141, 153, 158, 162, 165–167, 171, 177, 181, 225
  dynamic, 20, 241
  error, 2, 3, 13, 14, 16, 17, 22–24, 26, 29, 34, 104, 105, 108, 111, 120, 121, 131, 148, 151, 164, 168, 212, 218, 219, 225, 228, 243, 244
  inaccuracy of, 25, 174, 240, 244
  indirect, 18, 19, 27, 100, 141–171, 191, 204, 216, 221–226, 240, 248
  individual estimation of inaccuracy of, 244
  of ionization current, 233–235
  limits of error, 27
  mass, 21, 27
  equation, 1, 2, 141, 142, 148–150, 153, 160, 162, 163, 165–166, 168, 170, 174, 191, 216, 222, 225, 226, 229, 233, 236
  multiple, 20, 24, 101–102, 128–135, 173, 181, 239, 240, 244, 245
  observations, 17, 19, 20, 24, 27, 28, 71, 75–77, 79–85, 87, 89–91, 93, 94, 98, 101, 107, 118, 119, 128, 129, 138, 151, 157, 170, 181, 182, 195, 201, 218, 232–235, 242, 247
  parameters of alternating current, 15
  parameters of random processes, 16
  preliminary measurements, 22, 101
  range, 34, 35, 37, 39, 41, 42, 120, 121, 174, 207, 209, 210, 214, 233
  repeatability, 3
  reproducibility, 3
  simultaneous, 19, 173–192
  single, 101–103, 119–128, 131, 137, 161, 167, 181, 204–208, 240, 244, 245, 248
  standards
    initial measuring devices, 8, 35, 43, 59

primary, 7, 8
secondary, 7
working, 7
static, 20, 24
supplementary, 21
technical, 1, 31, 36
uncertainty, 11, 136, 243–248
unity, 3, 10
universal estimation of inaccuracy of, 21, 22, 212–213, 219
of voltage, 22, 118, 127, 168, 209, 216
vector, 144, 147, 152–154, 222, 225
of weak currents, 233
Measuring instruments
accuracy classes of measuring instruments, 46
analog, 32–34, 39, 56, 57, 59, 78, 84, 102, 103, 108, 110, 117, 131, 171, 179, 240
class index, 47
comparators, 31
dead band, 35, 43, 217
digital, 20, 23, 32, 33, 35, 41, 42, 45, 46, 103, 108, 121, 127, 215, 240
discrimination threshold, 35
drift, 8, 36, 45
indicating instruments, 20, 31, 32, 34, 39, 52, 105
instability, 36, 37, 45, 68, 130
material measures, 31
measurement channel, 33, 54, 111
measuring systems, 31, 47, 54, 111
measuring transducers, 7, 31, 32, 38, 52, 188–192
metrological characteristics, 33–37
nominal metrological characteristics, 36
readout devices, 35, 41, 42, 45
recording, 20, 24, 25, 27, 31, 33, 52, 102, 103, 240
scaling measuring transducers, 32
sensitivity, 31, 35, 48, 86, 89, 129, 217
serviceable, 54–58, 68
stability of measuring instruments, 69
uniformity of measuring instruments, 3, 8
unserviceable, 54, 56–58
value of scale division, 35
working measuring instrument, 7
Method of least squares, 173–175, 180–187
Method of maximum likelihood, 76, 100
Method of minimizing the sum of absolute deviations, 173
Method of moments, 73–75, 103
Method of substitutions, 180, 181
Methodological errors, 23, 105, 201
Methods of measurements
Borda's method, 117
Gauss's method of weighing, 118
method of contraposition, 118
method of replacement, 117
sign method of error compensation, 118
Methods of measurement data processing
method of reduction, 141, 152–154, 171, 191, 192, 225–226
method of transformation, 154–160
traditional method, 229–233
Metric Convention (Treaty of the Meter), 4, 11
Metrological characteristics, 33–36
Metrology
applied metrology, 8
general (theoretical) metrology, 8
legislative metrology, 8
particular metrology, 171
Model of object
parameter of a model, 14, 15
mathematical models, 16
Moments of random quantities, 73
Moving-coil galvanometer, 49

N

Nonlinear conditional equations, 180–181
Nonparametric methods, 92, 98, 221
Normal equations, 176, 177, 179, 183, 185
Normalized Gaussian function, 249

O

Object, 1, 2, 4, 10, 12–20, 22, 23, 31, 48, 81, 86, 87, 89, 101, 102, 105, 108, 109, 120, 124, 128, 161, 162, 209, 242
Operating conditions
rated, 37, 42, 44, 111, 122
reference, 23, 37, 39, 41, 44, 46, 111, 119, 122–123, 127, 131, 165, 166, 168, 207, 209, 222, 244, 246
Ordered series, 80, 93, 95, 97, 98
Output signal, 20, 34–36, 38, 43, 49, 51, 52, 54, 191, 241

P

Parameters of random process, 16
Pearson's test (χ^2 test), 87
Personal errors, 23
Physical constant, adjustment, 9
Potentiometer, 22, 106, 118, 215–220
Presentation of measurement results, 26–29
Primary measurement standard, 7, 8, 59, 64

Probability
of false rejection, 89
of false retention, 54, 55, 88

Q

Quantiles, 84, 85, 91, 93, 94, 98, 113, 114, 125, 131, 133–135, 149, 158, 170, 203, 231, 238, 247, 250
Quantity
active, 17, 18, 39
density of solid body, 12, 18, 142, 226–233
electrical resistance, 142, 144, 222–226
international system of units (SI), 6
passive, 17, 18, 39
units of, 6, 8
Quasirandom errors, 108–110

R

Random errors, 3, 23, 24, 27, 35, 39, 71, 83, 89, 102, 104–106, 108–110, 117, 124, 130–132, 136–139, 146–149, 157–159, 162, 165, 199–249
Random quantities, 16, 17, 48, 71–76, 78–80, 83, 85–88, 92, 93, 99, 109–111, 114, 129, 130, 132, 136, 143, 144, 150, 156, 164, 169, 180, 195, 201, 204, 205, 241
Reference conditions, 23, 37, 41, 44, 46, 111, 119–122, 127, 131, 165, 166, 168, 207, 209, 222, 244, 246
Relative error, 2, 3, 37, 39–43, 46, 48, 119, 121, 122, 124, 133, 156, 162, 163, 198, 199, 210, 213, 219, 222, 243
Relative frequency, 41, 222
Reproduction of the unit, 6, 8
Residuals, 104, 175, 177, 179, 180, 183, 186, 187, 190, 241
Robust estimates
truncated means, 95
weighted means, 95, 96
Winsor's means, 95, 96
Rules of rounding, 29

S

Second mixed moment, 143
Siegel–Tukey test, 93, 94
Significance level, 27, 66, 67, 81, 82, 85–94, 148, 224
Skewness, 75
Span, 40, 209
Square-root sum formula, 149, 151, 152, 228, 232
Standard cell, 118, 168, 215–219
Standard deviation, 9, 35, 80, 82–86, 108, 110, 111, 114, 129–132, 134, 148, 165, 179, 187, 202, 203, 205, 206, 231, 236
estimate of, 152, 160, 178, 196, 221, 224, 235
GUM, 246, 247
inaccuracy, 240
of random quantity, 74, 79
relative variance, 227, 232
uncertainty, 244
variance and, 158, 226
weighted mean, 237–238
Standardization of measuring instruments, 43, 87
Statistical tolerance interval, 86, 87
Strip strain gauges, 48
Student's distribution, 85, 91, 130, 131, 136–138, 148, 149, 151–153, 158, 159, 177, 184, 190, 191, 201, 203, 221, 227, 228, 231, 238, 247, 250

T

Taylor series, 145, 146, 149, 150, 158, 163, 168, 180, 184, 185, 228
Threshold discrepancy, 13–15, 23, 105, 161
Tolerance range, 86
Total uncertainty, 161–165
Traditional method of measurement data processing, 222–225
Transducer, 7, 31, 32, 34, 38, 40, 42, 52, 54, 61, 111, 188–192
Transfer function, 20, 34, 38, 49, 50, 54, 188–192
True value, 2, 3, 12–17, 21, 22, 34, 38, 39, 54, 55, 75, 76, 82–84, 86, 98, 104, 112, 125, 128, 129, 134, 141, 145, 146, 148, 151, 154, 162, 169, 177, 199, 206, 242–245, 247
Truncated means, 95
Type A uncertainty, 246
Type B uncertainty, 246

U

Unbiased estimate, 75, 78, 90, 153
Uncertainty, 133, 137, 158–160, 167, 171, 190–192, 197, 200, 202–205, 210, 213, 215, 217, 221, 222, 225–230, 232, 237, 238
confidence intervals, 2, 3
estimate of, 80, 152
indirect measurements, 161–165

intervallic estimates, 23
limits of error, 35, 45, 101, 103, 142,
measurement, 11, 22, 26, 27, 108–110, 125, 126, 131, 136, 139, 148, 149, 153–155, 241–248
Uniformity of measuring instruments, 3, 8
Units of basic quantities, 6, 8
Unity of measurements, 3, 10

V

Variable quantity, 17
Variance
random quantity, 143
weighted mean, 200, 237
Vector, measurement, 144, 147, 152–154, 222, 225
Verification, 14, 53, 116, 169, 171, 199, 206
Voltage divider, 105, 168, 215–219

W

Weighted mean, 95, 96, 159, 194, 196–206, 208, 231, 236, 237
Weights, 4, 19, 31, 51, 68, 96, 104–106, 108, 117, 118, 131, 159, 178, 179, 194–199, 201–205, 207, 208, 220, 226, 237
Welch–Satterthwaite formula, 148
Wilcoxon test, 92, 94
Winsor's mean, 95, 96
Working standards, 53, 54, 56, 58–61