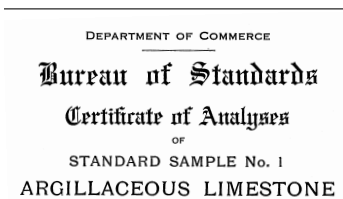


NIST Special Publication 260-202

**Evaluating, Expressing, and Propagating
Measurement Uncertainty for NIST
Reference Materials**



Antonio Possolo

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.260-202>

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Special Publication 260-202

**Evaluating, Expressing, and Propagating
Measurement Uncertainty for NIST
Reference Materials**

Antonio Possolo
*Statistical Engineering Division
Information Technology Laboratory*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.260-202>

July 2020



U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

National Institute of Standards and Technology Special Publication 260-202
Natl. Inst. Stand. Technol. Spec. Publ. 260-202, 63 pages (July 2020)
CODEN: NSPUE2

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.260-202>

Foreword

This document is intended for users of NIST reference materials who wish to gain some understanding of how NIST evaluates and expresses measurement uncertainty reported in certificates of NIST reference materials, and also for users who aim to learn how measurement uncertainty may be propagated to derivative quantities that involve the corresponding measured values. The document thus serves as a supplement and companion of NIST Special Publication 260-136 (2020 Edition).

The table of contents, the lists of examples and of reference materials, as well as the orientation and road map presented in the first few pages, include navigation aids, in the form of hyperlinks, that enable direct and easy access to examples illustrating different ways of expressing or propagating uncertainty, and to specific reference materials that are mentioned in this document.

The document reviews the concepts of measurement, measurement uncertainty, and reference material, and includes a concise refresher of concepts and devices from probability and statistics that are often used in certificates of NIST reference materials.

The bulk of the document comprises specific descriptions and concrete examples of how NIST evaluates measurement uncertainty for NIST reference materials, of how this uncertainty is expressed, and of how it may be propagated in applications that use NIST reference materials. Illustrative computer codes implementing calculations described in some of the examples are included in an appendix.

Keywords

amino acid; amount fraction; approximation; asymmetric; beta; bottom-up; calibration; categorical; central limit theorem; coverage factor; coverage probability; degrees of freedom; DNA; expression; evaluation; expanded uncertainty; functional measurand; gas mixture; Gaussian; lognormal; mass fraction; measurand; measurement; measurement equation; functional; measurement model; measurement uncertainty; nominal; nucleobase; observation equation; ordinal; paramagnetic; probability; propagation; protein; qualitative measurand; quantification; quantitative measurand; reference material; skew-normal; standard reference material; standard uncertainty; statistical model; statistical test; statistics; Student; type A; type B; top-down; welch-satterthwaite.

Typesetting

The cover and title pages, as well as this page, are typeset in Adobe Times Roman. The rest of this document is typeset in Linux Libertine, a digital typeface created by the Libertine Open Fonts Project, together with New TX math fonts, and it was composed using \LaTeX as implemented in the 2020 TeX Live distribution from the TeX Users Group.

Table of Contents

EVERY ITEM IN THIS TABLE, EXCEPT THE PAGE NUMBERS, IS A HYPERLINK

1	Orientation	8
2	Road Map	9
3	Notation	10
4	Introduction	11
	Measurement result	11
	Measurement	11
	Measurement uncertainty	12
5	Statistical and Probabilistic Concepts	13
	Standard Uncertainty, Expanded Uncertainty	13
	Chance and Probability	13
	Probability distribution, Random Variable	14
	Student's <i>t</i> distribution	14
	Gaussian distribution	16
	Lognormal distribution	16
	Skew-normal distribution	17
	Uniform (Rectangular) distribution	17
	Beta distribution	17
	Linear combinations of uncorrelated random variables	17
	Welch-Satterthwaite	17
6	Evaluating Measurement Uncertainty	20
	Measurement model	20
	Measurement uncertainty: Bottom-up evaluation	20
	EXAMPLE 6.3.A: Uncertainty budget supporting the gravimetric determination of the mass fraction of mercury in NIST SRM 1641e Mercury in Water	20
	Measurement uncertainty: Type A and Type B evaluations	21
	Measurement uncertainty: Top-down evaluation	21
	EXAMPLE 6.6.A: Type A and Type B evaluations of contributions from sources of uncertainty, in the determination of the mass fraction of arsenous acid (H ₃ AsO ₃) in NIST SRM 3037 Arsenous Acid (AsIII) Standard Solution	21
	Nominal and ordinal properties	21
	EXAMPLE 6.7.A: Species identification in NIST SRM 3246 <i>Ginkgo biloba</i> (Leaves) using sequences of nucleobases in the plant's DNA	22
	EXAMPLE 6.7.B: Sequence of amino acids in NIST SRM 927e Bovine Serum Albumin (7 % Solution), and its relative molecular mass	22

EXAMPLE 6.7.C: Quantification of the uncertainty associated with identifications of nucleobases in a DNA strand, based on quality scores produced by base-calling software <i>Phred</i>	22
7 Expressing Measurement Uncertainty	24
Coverage interval	24
EXAMPLE 7.2.A: Comparing a user's determination of the mass concentration of 25-hydroxyvitamin D ₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum, (High Level), with the certified value	24
EXAMPLE 7.2.B: Computing the number of degrees of freedom implied by the coverage factor used to produce the expanded uncertainty of the mass fraction of ascorbic acid (vitamin C) in NIST SRM 3280 Multivitamin/Multielement Tablets	25
EXAMPLE 7.2.C: Evaluating between-method differences (<i>dark uncertainty</i>) in the measurement of the mass fraction of tin in NIST SRM 3161a Tin (Sn) Standard Solution	25
EXAMPLE 7.2.D: Comparing determinations of the mass concentration of albumin in NIST SRM 927e Bovine Serum Albumin (7 % Solution) made using a novel analytical method, with the certified value	26
Asymmetrical coverage interval	27
EXAMPLE 7.3.A: Comparing determinations of the mass fraction of chromium in two reference materials, NIST SRM 606 Trace Elements in Basalt Glass and NIST SRM 688 Basalt Rock, which were prepared starting from the same raw material	27
EXAMPLE 7.3.B: Asymmetric coverage interval for the mass fraction of chromium in NIST SRM 2780a Hard Rock Mine Waste	29
EXAMPLE 7.3.C: Asymmetric coverage interval for the purity of benzoic acid in NIST PS1 Primary Standard for quantitative NMR (Benzoic Acid)	29
Functional measurands	29
EXAMPLE 7.4.A: Comparing particle size distributions for NIST SRM 1003c Glass Beads (Particle Size Distribution) that were characterized using two different measurement methods	29
EXAMPLE 7.4.B: Value assignment and uncertainty evaluation for the amount fraction of oxygen in nitrogen in NIST SRM 2657a Oxygen in Nitrogen	31
8 Propagating Measurement Uncertainty	32
Older reference materials	32
EXAMPLE 8.1.A: Using the <i>NIST Uncertainty Machine</i> to characterize the uncertainty of the ratio of the mass fractions of lead and uranium in NIST SRM 278 Obsidian Rock	32

EXAMPLE 8.2.A: Computing the uncertainty associated with the logarithm of the mass fraction of chromium in NIST SRM 3112a Chromium (Cr) Standard Solution	33
EXAMPLE 8.3.A: Evaluating the uncertainty of the volume of silver nanoparticles in NIST RM 8017 Polyvinylpyrrolidone Coated Silver Nanoparticles, assumed to be spherical, whose diameters were measured using transmission electron microscopy	34
Products of powers of input quantities	34
EXAMPLE 8.4.A: Evaluating uncertainty associated with amount-of-substance concentration of 25-hydroxyvitamin D ₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)	35
Propagating uncertainty associated with a functional measurand	37
EXAMPLE 8.5.A: Using NIST SRM 1003c Glass Beads (Particle Size Distribution) to calibrate measurements of particle size	37
Propagating uncertainties expressed asymmetrically	39
EXAMPLE 8.8.A: Estimating and evaluating the uncertainty associated with the ratio $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$ in NIST SRM 606 Trace Elements in Basalt Glass	39
Propagating uncertainty for qualitative measurands	41
EXAMPLE 8.9.A: Propagating the uncertainty surrounding identification of nucleobases in a DNA strand, to the Damerau-Levenshtein distance between a measured sequence and a reference sequence	41
Propagating probability distributions	42
EXAMPLE 8.10.A: Characterizing the individual and joint uncertainty of two ratios of elemental amount fractions used in Pearce Element Ratio analysis in geochemistry, in NIST SRM 2780a Hard Rock Mine Waste	42
References	45
Appendix: R Code	52
R environment for statistical computing and graphics	52
EXAMPLE 7.2.A: NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)	52
EXAMPLE 7.2.B: NIST SRM 3280 Multivitamin/Multielement Tablets	52
EXAMPLE 7.2.C: NIST SRM 3161a Tin (Sn) Standard Solution	53
EXAMPLE 7.2.D: NIST SRM 927e Bovine Serum Albumin (7 % Solution)	53
EXAMPLE 7.3.A: NIST SRM 606 Trace Elements in Basalt Glass	54
EXAMPLE 7.4.A: NIST SRM 1003c Glass Beads (Particle Size Distribution)	55
EXAMPLE 8.4.A: NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)	55
EXAMPLE 8.5.A: NIST SRM 1003c Glass Beads (Particle Size Distribution)	56
EXAMPLE 8.8.A: NIST SRM 606 Trace elements in basalt	57

List of Examples

- 6.3.A Uncertainty budget supporting the gravimetric determination of the mass fraction of mercury in NIST SRM 1641e Mercury in Water
- 6.6.A Type A and Type B evaluations of contributions from sources of uncertainty, in the determination of the mass fraction of arsenous acid (H_3AsO_3) in NIST SRM 3037 Arsenous Acid (AsIII) Standard Solution
- 6.7.A Species identification in NIST SRM 3246 *Ginkgo biloba* (Leaves) using sequences of nucleobases in the plant's DNA
- 6.7.B Sequence of amino acids in NIST SRM 927e Bovine Serum Albumin (7 % Solution), and its relative molecular mass
- 6.7.C Quantification of the uncertainty associated with identifications of nucleobases in a DNA strand, based on quality scores produced by base-calling software *Phred*
- 7.2.A Comparing a user's determination of the mass concentration of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum, (High Level), with the certified value
- 7.2.B Computing the number of degrees of freedom implied by the coverage factor used to produce the expanded uncertainty of the mass fraction of ascorbic acid (vitamin C) in NIST SRM 3280 Multivitamin/Multielement Tablets
- 7.2.C Evaluating between-method differences (*dark uncertainty*) in the measurement of the mass fraction of tin in NIST SRM 3161a Tin (Sn) Standard Solution (Lot No. 140917)
- 7.2.D Comparing determinations of the mass concentration of albumin in NIST SRM 927e Bovine Serum Albumin (7 % Solution) made using a novel analytical method, with the certified value
- 7.3.A Comparing determinations of the mass fraction of chromium in two reference materials, NIST SRM 606 Trace Elements in Basalt Glass and NIST SRM 688 Basalt Rock, which were prepared starting from the same raw material
- 7.3.B Asymmetric coverage interval for the mass fraction of chromium in NIST SRM 2780a Hard Rock Mine Waste
- 7.3.C Asymmetric coverage interval for the purity of benzoic acid in NIST PS1 Primary Standard for quantitative NMR (Benzoic Acid)
- 7.4.A Comparing particle size distributions for NIST SRM 1003c Glass Beads (Particle Size Distribution) that were characterized using two different measurement methods
- 7.4.B Value assignment and uncertainty evaluation for the amount fraction of oxygen in nitrogen in NIST SRM 2657a Oxygen in Nitrogen
- 8.1.A Using the *NIST Uncertainty Machine* to characterize the uncertainty of the ratio of the mass fractions of lead and uranium in NIST SRM 278 Obsidian Rock
- 8.2.A Computing the uncertainty associated with the logarithm of the mass fraction of chromium in NIST SRM 3112a Chromium (Cr) Standard Solution (Lot No. 030730)
- 8.3.A Evaluating the uncertainty of the volume of silver nanoparticles in NIST RM 8017 Polyvinylpyrrolidone Coated Silver Nanoparticles, assumed to be spherical, whose diameters were measured using transmission electron microscopy
- 8.4.A Evaluating uncertainty associated with amount-of-substance concentration of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)
- 8.5.A Using NIST SRM 1003c Glass Beads (Particle Size Distribution) to calibrate measurements of particle size
- 8.8.A Estimating and evaluating the uncertainty associated with the ratio $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$ in NIST SRM 606 Trace Elements in Basalt Glass
- 8.9.A Propagating the uncertainty surrounding identification of nucleobases in a DNA strand, to the Damerau-Levenshtein distance between a measured sequence and a reference sequence
- 8.10.A Characterizing the individual and joint uncertainty of two ratios of elemental amount fractions used in Pearce Element Ratio analysis in geochemistry, in NIST SRM 2780a Hard Rock Mine Waste

List of Reference Materials

The (blue) linking references point either to paragraphs where a listed reference material is mentioned, or to examples illustrating use of information provided in the respective certificate.

SRM 278	Obsidian Rock	8.1.A
SRM 606	Trace Elements in Basalt Glass	7.3.A , 8.8.A
SRM 688	Basalt Rock	7.3.A , 8.8.A
SRM 916a	Bilirubin	6.2
SRM 927e	Bovine Serum Albumin (7 % Solution)	6.7.B , 7.2.D
SRM 1003c	Glass Beads (Particle Size Distribution)	7.4 , 7.4.A , 8.5 , 8.5.A
SRM 1641e	Mercury in Water	6.3.A
SRM 1720	Northern Continental Air	6.2
SRM 2034	Holmium Oxide Solution Wavelength Standard (240 nm to 650 nm)	7.4
SRM 2242a	Relative Intensity Correction Standard for Raman Spectroscopy (532 nm Excitation)	7.4
SRM 2374	DNA Sequence Library for External RNA Controls	4.2
SRM 2657a	Oxygen in Nitrogen	7.4.B
SRM 2780a	Hard Rock Mine Waste	4.3 , 7.3.B , 8.10.A
SRM 2973	Vitamin D Metabolites in Frozen Human Serum (High Level)	4.2 , 7.2.A , 8.4.A
SRM 3037	Arsenous Acid (AsIII) Standard Solution	6.6.A
SRM 3112a	Chromium (Cr) Standard Solution (Lot No. 030730)	8.2.A
SRM 3156	Tellurium (Te) Standard Solution (Lot No. 140830)	5.5
SRM 3161a	Tin (Sn) Standard Solution (Lot No. 140917)	7.2.C
SRM 3246	<i>Ginkgo biloba</i> (Leaves)	4.3 , 6.7.A
SRM 3280	Multivitamin/Multielement Tablets	7.2.B
RM 8017	Polyvinylpyrrolidone Coated Silver Nanoparticles	8.3.A
CRM PS1	Primary Standard for quantitative NMR (Benzoic Acid)	7.3.C

Since the evaluation and expression of uncertainty for NIST SRMs supporting Charpy Impact Machine testing are thoroughly documented in separate NIST publications [[Splett et al., 2008a,b](#)], such SRMs are not included as examples in the present one.

Notices and Disclaimers

The certificates of NIST reference materials are not listed under References because they are all readily accessible at <https://www.nist.gov/srm>, except for NIST CRM PS1, whose certificate is available upon request from the NIST Material Measurement Laboratory.

The term “certificate” is used throughout as shorthand for the document that accompanies each reference material produced by NIST, regardless of whether the values assigned to the material properties are certified or non-certified. Similarly, the term “reference material” encompasses all types of reference materials, and the term “reference value” likewise means any value assigned to a material property in a manner that satisfies the requirements specified by [Beauchamp et al. \[2020\]](#).

Several examples in this document involve measurement results listed in certificates of actual NIST reference materials, but may combine them with other, fictional data for pedagogic reasons. If any results in these examples, or results that may be derived from them, should differ from results stated in any certificates of NIST reference materials, then they shall be regarded merely as illustrative examples that do neither impugn nor replace the published reference values or associated uncertainties.

In all cases, the most recent versions of certificates published and maintained by NIST, and companion NIST Special Publications associated with specific materials, are the sole authoritative sources of certified (and non-certified) values, their associated uncertainties, and other related information.

Computations described and illustrated in this document could be carried out using any software environment where statistical calculations are implemented reliably. Here they are all done using the R environment for statistical computing and graphics [[R Core Team, 2020](#)].

The [Appendix](#) describes how to obtain, install, and use R, which is freely available for all major computer operating systems, and lists R codes for computations supporting several of the examples. These codes are offered without warranty of any kind concerning their suitability for any particular or general purpose.

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Acknowledgments

The author is much indebted to the following NIST colleagues for their detailed, rigorous, perspicacious, generous reviews of drafts of this document. Their questions, comments, suggestions, and corrections enabled many, very significant improvements, and also greatly enriched the author's understanding of the substantive issues involved:

Ashley Beasley-Green	Carlos Beauchamp	Johanna Camara	Christina Cecelski
Steve Choquette	Paul DeRose	David Duewer	Vincent Hackley
Katrice Lippa	Enrico Lucon	John Molloy	Michael Nelson
Melissa Phillips	Katherine Sharpless	John Sieber	Jolene Splett
Michael Winchester	Donald Windover	Justin Zook	

The author is grateful to Kathryn Miller (Information Services Office) for production guidance and assistance.

1 Orientation

This document is a companion of NIST Special Publication 260-136 (2020 Edition) [[Beauchamp et al., 2020](#)], that explains how NIST expresses measurement uncertainty associated with reference values, and exemplifies how such uncertainty may be propagated to derivative quantities that involve these values.

The **Road Map** for navigating this document (**Section 2**) provides direct hyperlinks to examples of NIST reference materials illustrating different ways of expressing or propagating uncertainty. Users wishing to understand only the specific information about measurement uncertainty provided in a particular certificate, may use this road map to jump to the location in this document where such information is explained.

Section 3 (Notation) provides a succinct overview of the notation used in certificates of NIST reference materials, and a brief introduction to the meaning of the corresponding symbols. The other sections provide more detailed accounts of notation and underlying concepts, as needed.

Section 4 (Introduction) provides basic information about NIST reference materials, and reviews the concepts of *measurement* and *measurement uncertainty*.

Section 5 (Statistical and Probabilistic Concepts) provides a refresher of a few concepts of statistics and probability that appear in certificates of reference materials, and that are used throughout this document.

Section 6 (Evaluating Measurement Uncertainty) explains how NIST evaluates measurement uncertainty.

Section 7 (Expressing Measurement Uncertainty) provides many examples illustrating different ways to express measurement uncertainty that are used in certificates of NIST reference materials.

Section 8 (Propagating Measurement Uncertainty) gathers practical guidance for propagating uncertainty, and illustrates it in concrete examples, several of which develop further some of the examples introduced in previous sections.

The **Appendix** provides illustrative R code that implements the calculations supporting several of the examples, and explains how the R environment for statistical computing and graphics can be installed.

2 Road Map

To reach guidance for how to propagate the uncertainty reported in the certificate of a NIST reference material, or to gain familiarity with how the uncertainty is expressed, click one of the colored (blue) example labels in the row of the following table that best describes the information the certificate provides:

INFORMATION	EXAMPLES
<i>u</i>	8.1.A , 8.4.A
<i>U</i>	6.6.A , 7.2.A , 7.2.C
<i>u</i> and <i>k</i>	8.4.A
<i>U</i> and <i>k</i>	7.2.A , 7.2.B , 8.2.A , 8.3.A
<i>U</i>, <i>k</i>, and ν	7.2.C , 8.4.A
Coverage Interval	7.3.A , 7.3.B , 7.3.C , 7.4.B , 8.8.A , 8.10.A
Asymmetric Coverage Interval	7.3.A , 7.3.B , 7.3.C , 8.8.A , 8.10.A
Replicated Determinations	7.2.D
Probability Distribution	7.3.B , 7.3.C , 7.4.A , 8.5.A , 8.10.A
Functional Measurand	7.4.A , 8.5.A
Qualitative Measurand	6.7.A , 6.7.C , 8.9.A

- ***u***, ***U***, ***k***, and **ν** are labels referring to standard uncertainty, expanded uncertainty, coverage factor, and number of degrees of freedom, respectively — their meaning is explained next, under [Notation](#);
- **Coverage Interval** means that the uncertainty is expressed as a coverage interval stated explicitly, which need not be centered at the measured value;
- **Probability Distribution** refers to cases where the uncertainty is expressed as a fully specified probability distribution, or as a large sample drawn from one such;
- **Functional Measurand** indicates that the measurand is a function (for example a spectrum), not a scalar;
- **Qualitative Measurand** points to examples where the measurand is a qualitative property (ordinal or nominal).

To return to the table above, click the label of any example throughout this document, because they all point back to this [Road Map](#).

Clicking on a colored (blue) bibliographic reference anywhere in this document, will take the reader to the list of References, beginning on Page [45](#). To get back to the page where the reference was made, click the colored (blue) number of the referring page that appears, in a smaller font size, between square brackets at the end of the description of the bibliographic item.

3 Notation

The *expanded uncertainty* $U_p(x)$ associated with a value x of a scalar quantity X is the most common expression of measurement uncertainty in certificates of NIST reference materials.

For example, the certificate for NIST SRM 3161a Tin (Sn) Standard Solution (Lot No. 140917) states the certified value as $10.011 \text{ mg/g} \pm 0.025 \text{ mg/g}$. The “0.025 mg/g” is an expanded uncertainty. The statement conveys the belief that the true value of the mass fraction of tin in the solution is somewhere between 9.986 mg/g and 10.036 mg/g, with 95 % probability.

The *coverage probability* $0 < p < 1$ (often called simply “probability” or “confidence”), usually written as a percentage (say, 95 %), quantifies the strength of the belief that the true value of the quantity lies within the interval $x \pm U_p(x)$.

Many certificates state the coverage probability separately, which typically is 95 %, and omit the subscript from the symbol for the expanded uncertainty, writing $U(x)$ instead, or even just U when the value it pertains to is clear from the context.

The corresponding *standard uncertainty* is denoted $u(x)$, or $u_c(x)$ when one wishes to emphasize that it is a combined standard uncertainty in the sense of [JCGM 100:2008](#), 5.1.1 (GUM), comprising contributions from multiple sources of uncertainty that will have been evaluated individually, each with its own standard uncertainty.

In many certificates where both the expanded and standard uncertainties are listed, $u(x) = U_p(x)/k_p$, where k_p is the *coverage factor*, whose subscript p is usually omitted. In many cases, k_p is a suitable percentile either from a Gaussian or from a Student’s t probability distribution.

The coverage factor may be meaningful only when the underlying probability distribution that describes measurement uncertainty is symmetric, and it depends not only on the coverage probability p but also on the nature of the underlying probability distribution that fully characterizes measurement uncertainty: whether this distribution is Gaussian, rectangular, etc.

Many certificates state explicitly the *number of degrees of freedom* that $u(x)$ and $U_p(x)$ are based on, or they may state k_p only. In some cases, $u(x)$ and $U_p(x)$ are computed one separately from the other. This is often the case when uncertainties are evaluated using Monte Carlo methods. In such cases, the number of degrees of freedom is only “virtual”, having been computed as the ratio $U_p(x)/u(x)$.

The meaning of the interval $x \pm U_p(x)$ depends neither on the number of degrees of freedom nor on the coverage factor. However, to propagate the uncertainty that it conveys, to derivative quantities involving x , additional information is required, or assumptions will have to be made if no such information is available.

4 Introduction

4.1 The development of reference materials is an essential and defining element of the mission of the National Institute of Standards and Technology (NIST). Several NIST laboratories produce reference materials, but the vast majority are produced by the Material Measurement Laboratory (MML), following the guidance in NIST Special Publication 260-136 (2020 Edition) [Beauchamp et al., 2020], under the [NIST Quality Management System](#), and consistent with applicable international standards.

A *reference material* (RM) is a material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process [Emons et al., 2006] [ISO, 2015, 2.1.1]. These properties can be quantitative or qualitative.

4.2 Each reference material is accompanied by a certificate that describes the nature of the material, its intended purpose, how it shall be used, and the expiration date for the certification. The certificate lists measurement results for one or more material properties, which may be quantitative or qualitative. The mass fraction of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level) is an example of the former, and sequences of nucleobases in DNA of NIST SRM 2374 DNA Sequence Library for External RNA Controls is an example of the latter.

4.3 In many cases, reference materials are documented in much greater detail than their certificates allow, in 260-series NIST Special Publications, for example [Paule and Mandel \[1970\]](#) or [Sharpless et al. \[2015\]](#). In some cases, for example for NIST SRM 3246 *Ginkgo biloba* (Leaves) and for NIST SRM 2780a Hard Rock Mine Waste, supplementary information about the reference materials is publicly available on the World Wide Web.

4.4 A measurement result comprises a measured value, and an evaluation of the associated measurement uncertainty performed consistently with the NIST Quality Manual [NIST, 2019, Appendix C] and with NIST Technical Notes 1297 [Taylor and Kuyatt, 1994] and 1900 [Possolo, 2015] (NIST *Simple Guide*). The evaluation of measurement uncertainty should also be consistent with the provisions of the *Guide to the expression of uncertainty in measurement* (GUM) [JCGM 100:2008].

4.5 NIST understands *measurement* to be an experimental or computational process that, by comparison with a standard of reference, produces an estimate of the true value of a property of a material or virtual object or collection of objects, or of a process, event, or series of events, together with an evaluation of the uncertainty associated with that estimate, and intended for use in support of decision-making [NIST, 2019, 3. Definitions]. The *measurand* (the property that is the object of measurement) may be quantitative (scalar, vectorial, or functional), or qualitative (either nominal or ordinal).

The “standard of reference” may be a reference material, a realization of an SI unit, or some measurement reference scale that a relevant community recognizes, for example for ratios (say, “chemical shifts” in nuclear magnetic

resonance, or delta values for isotopic compositions), counts (say, of electrons in coulometry, or of replicates resulting from amplification in digital PCR), etc.

4.6 *Measurement uncertainty* is the doubt about the true value of the measurand that remains after making a measurement.

Measurement uncertainty can be quantified both for quantitative and for qualitative measurands, or it may be expressed only qualitatively: for example, asserting with highest confidence that a particular leaf is from a plant of the species *Ginkgo biloba*, or merely indicating that a particular Portland cement clinker is very likely to contain more than 60 cg/g of alite.

The most complete quantitative description of measurement uncertainty is in terms of a fully specified probability distribution on the set of values of the measurand, which can be done both for quantitative and for qualitative measurands. Example 7.3.B describes an instance of the former, and Example 6.7.C describes an instance of the latter.

In most cases, measurement uncertainty is expressed summarily and approximately by specifying only a coverage interval, or a particular indication of the dispersion (or scatter) of such distribution, for example an expanded uncertainty or the standard uncertainty [NIST, 2019, 3. Definitions].

5 Statistical and Probabilistic Concepts

This section reviews several concepts and devices from probability and statistics that are used either in certificates of NIST reference materials or in this document. These technicalities, which many users of our reference materials find daunting, are necessitated by the fact that the evaluation, expression, and interpretation of measurement uncertainty involve those concepts and devices.

[Freedman et al. \[2007\]](#) provide a leisurely, entertaining, accessible introduction to basic concepts and techniques of probability and statistics, requiring no more than high-school algebra as mathematical prerequisite, and indeed sufficing as preparation for the reader to benefit from this document. The more recent account by [Diez et al. \[2019\]](#) is a well regarded, freely downloadable alternative.

[Hodges and Lehmann \[2005\]](#) is less popular than either of those two, more focused on the technicalities but still with the same minimal prerequisites, and was written by two luminaries in the field.

Readers who will have studied calculus in high-school or college, may like to refer to the 2nd edition of Morris DeGroot's (1986) *Probability and Statistics* [[DeGroot, 1986](#)], which is out of print but may still be found as a used book, or to [Possolo and Toman \[2011\]](#)'s tutorial on the same topics, which is intended specifically for metrologists.

5.1 STANDARD UNCERTAINTY. The standard uncertainty u is the standard deviation of the probability distribution used to describe measurement uncertainty (the concept of *probability distribution* is reviewed below), or a quantity with a similar meaning: it has the same units as the measurand.

The meaning of the standard uncertainty depends on the underlying probability distribution. For example, the probability of obtaining a value within one standard deviation of the mean is 68 % for a Gaussian (or normal) distribution, but it is 57 % for a uniform (or rectangular) distribution, and 82 % for a Student's t distribution with 3 degrees of freedom.

5.2 EXPANDED UNCERTAINTY. The expanded uncertainty, U_p , for a specified coverage $0 < p < 1$, is an expression of measurement uncertainty such that, with probability (or confidence) p , the true value of the measurand is believed to lie within U_p of the measured value. The expanded uncertainties reported in NIST certificates are for 95 % coverage probability (that is, $p = 0.95$). The ratio between expanded and standard uncertainties is the *coverage factor* $k_p = U_p/u$.

5.3 CHANCE AND PROBABILITY. Probability is a concept that may be interpreted in many different ways [[Hájek, 2007](#)]. The interpretation that is most apt for use in the quantification of measurement uncertainty is as expression of degree of belief (in the truth of an assertion about the measurand) reflecting the extent of one's knowledge about the true value of the measurand [[O'Hagan, 2014](#)].

[Campbell \[1920, Chapter VII\]](#) distinguishes *chance* from *probability*, employing the former for events that show no regularity whatever and for which no forecast can be

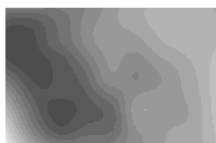
made, and using the latter as a measure of one's degree of knowledge.

Carnap [1962, Chapter II] also favors a dual understanding, distinguishing two kinds of probability. One kind expresses degree of confirmation of a hypothesis with respect to a corpus of evidence. The other kind describes a property ascribed to unpredictable events, as in games of chance, for example roulette, or in radioactive disintegrations.

It should be noted that even such “physical” probabilities as seem to be intrinsic properties of games of chance and radionuclides, need not necessarily be interpreted as frequencies in the long run, for they can be regarded as expressions of *propensity* instead [Popper, 1990].

Frequentist interpretations of probability, either in finite [Hájek, 1996] or in infinite sequences of events [Hájek, 2009], are notoriously deficient. However, this does not necessarily undermine the practical value of probabilistic or statistical methods based on interpretations of probability as hypothetical frequencies in the long run [La Caze, 2016].

5.4 PROBABILITY DISTRIBUTION. Imagine an explorer looking for buried treasure with the aid of a map that is colored in shades of gray: the darker the shade over a region, the more likely it is for treasure to be buried there.



A *probability distribution* is like these shades of gray, or like a distribution of mass over the set of possible values for a measurand: the true value of the measurand is more likely to be where the shades are darkest, or where the mass density is the largest.

5.5 RANDOM VARIABLE. To characterize the uncertainty surrounding an estimate of a (quantitative or qualitative) property — say, the mass fraction of tellurium in NIST SRM 3156 Tellurium (Te) Standard Solution (Lot No. 030730) — it is convenient to model it as a *random variable*. This does not mean that the property is changeable, or that it varies unpredictably. It means simply that the property in question has a probability distribution as an attribute that describes how well (or how poorly) one knows its true value.

For example, the certificate of NIST SRM 3156 lists the certified value of the mass fraction of tellurium in the solution as 10.005 mg/g, and states that NIST is 95 % confident that the corresponding true value lies between 9.967 mg/g and 10.043 mg/g.

The certificate provides additional information stating that $k_p = 2.165$, and that U_p is based on 12.7 degrees of freedom. The additional information suggests, albeit implicitly, that NIST assigns a very specific probability distribution as attribute of that mass fraction, as means to characterize the associated uncertainty. The specific distribution is a (rescaled and shifted) Student's t distribution with 12.7 degrees of freedom, mean 10.011 mg/g, and standard deviation $\frac{1}{2}(10.043 \text{ mg/g} - 9.967 \text{ mg/g})/2.165 = 0.01755 \text{ mg/g}$.

5.6 STUDENT'S t DISTRIBUTION. The (rescaled and shifted) Student's t probability distribution has three defining parameters: median μ , scale $\tau > 0$, and number of degrees

of freedom $\nu > 0$. The distribution is symmetrical around μ , and its probability density is bell-shaped.

The Gaussian distribution (5.7) also has a bell-shaped probability density. However, if one compares the shapes of the two distributions, when they have the same median and the same scale, Student's has the heavier tails — meaning that the probability of large deviations from the median is greater for Student's than for the Gaussian. The smaller the number of degrees of freedom ν , the heavier the Student's t tails by comparison with the tails of a Gaussian distribution with the same median and scale [DeGroot and Schervish, 2012]. If $\nu > 2$, then the distribution has finite standard deviation $\tau\sqrt{\nu/(\nu-2)}$. The larger the ν , the closer to Gaussian the Student's t distribution becomes.

The Student's t distribution goes hand in hand with the Gaussian distribution, owing to the following fact concerning the average \bar{x} and standard deviation s of a sample x_1, \dots, x_m drawn from a Gaussian distribution with mean μ and standard deviation σ (both unknown): $(\bar{x} - \mu)/(s/\sqrt{m})$ is like an outcome of a Student's t random variable with median 0, scale 1, and $m - 1$ degrees of freedom. The Student's t distribution thus gets around the difficulty caused by not knowing σ , and makes due allowance for the fact that s is not σ [Mosteller and Tukey, 1977].

The reason why the GUM (particularly in its Annex G) gives such pride of place to the Student's t distribution is a little more convoluted than the relationship between this distribution and averages and standard deviations of Gaussian samples reviewed above.

The GUM is mostly concerned with output quantities that, in small neighborhoods of their true values, are approximately linear functions of the true values of the corresponding input quantities. On the one hand, this ensures that the formula the GUM offers for $u_c(y)$ is accurate.

On the other hand, that approximation implies that the output quantity Y is approximately equal to a sum of independent random variables (if the input quantities indeed can be so modeled). The GUM G.2 then invokes a result from probability theory, the so-called *Central Limit Theorem* [DeGroot and Schervish, 2012, Theorem 6.3.3], to argue that, under specified conditions, Y will have a probability distribution that is approximately Gaussian.

In these circumstances, if η denotes the true value of Y , then $(y - \eta)/u_c(y)$ is approximately like an outcome of a random variable with a Student's t distribution whose number of degrees of freedom is given approximately by the Welch-Satterthwaite formula (reviewed below, in 5.13). This is the reason why the GUM suggests using coverage factors that are suitable percentiles of a Student's t distribution.

To compute the probability that the true mass fraction of tellurium in NIST SRM 3156 (Lot No. 030730) lies in a specified interval, say between 10.000 mg/g and 10.050 mg/g, using a Student's t distribution, execute the following lines of R code:

```
mu = 10.005; U = 0.038
k = 2.165; nu = 12.7
u = U / k
Lwr = (10.000 - mu) / u
Upr = (10.050 - mu) / u
pt(Upr, df=nu) - pt(Lwr, df=nu)
```

This R code computes the area under the probability density curve of a Student's t distribution with 12.7 degrees of freedom that lies between the standardized endpoints

of the specified interval. The standardized endpoints are $(10.000 - 10.005)/(0.038/2.165)$ and $(10.050 - 10.005)/(0.038/2.165)$.

Once the coverage probability p has been chosen, the coverage factor k_p and the number of degrees of freedom ν are not independent, but one determines the other, as explained in Example 7.2.B and developed in A.3. The number of degrees of freedom typically refers to the number of independent pieces of information that contribute to an estimate of a standard deviation, in which case one expects it to be an integer. However, and in general, the number of degrees of freedom of the Student's t distribution is just a parameter that serves to distinguish different members of this family from one another, and need not be an integer (but must be positive).

5.7 GAUSSIAN DISTRIBUTION. The Gaussian distribution is usually called “normal” distribution. Since Carl Friedrich Gauss [Gauss, 1809] was the first to explain, in 1809, the role of this distribution in the characterization of measurement errors, and also because John Tukey for good reason casts doubt on the “normality” of the normal distribution [Tukey, 1977], we call it “Gaussian” throughout.

The certified mass concentration of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level) is 39.4 ng/mL, and the expanded uncertainty is 0.8 ng/mL. The certificate states that the coverage factor was $k = 2$.

In practice, a coverage factor of 2 is often regarded as the result of rounding 1.96, which is the factor for 95 % coverage when the output quantity is Gaussian. Proceeding on this assumption, the probability that the true mass fraction is greater than 40.0 ng/mL, computed using the R language [R Core Team, 2020], is $1 - \text{pnorm}(40, \text{mean}=39.4, \text{sd}=0.8/2) \approx 7\%$.

Alternatively, one may interpret all coverage factors as percentiles of Student's t distributions, and compute the corresponding degrees of freedom using the R function `kInverter`, defined in A.3 and used in Example 7.2.B. In Example 8.3.A the coverage factor 2 is interpreted as the 97.5th percentile of a Student's t distribution with 60.4 degrees of freedom, which for just about all practical purposes is interchangeable with a Gaussian distribution.

There are other probability distributions (some not even symmetrical) for which the interval centered at the mean, of half-width 2 standard deviations, does encompass approximately 95 % of the corresponding probability [Freedman et al., 2007, Page 81]. However, there is no mathematical assurance that it will be so in general.

5.8 LOGNORMAL DISTRIBUTION. If a positive quantity X is such that $\log X$ has a Gaussian distribution, then X is said to have a lognormal distribution. In this document, “log” refers to the natural logarithm, that is, the logarithm base e , so that $x = \exp(\log(x))$ for every $x > 0$. Logarithms base 10 are denoted \log_{10} , as in Example 8.2.A.

If the relative uncertainty modeled by this distribution is small (say, less than 10 %), then the lognormal distribution is almost indistinguishable from a Gaussian distribution with the same mean and standard deviation.

This fact can help avoid difficulties when applying the Monte Carlo method for un-

certainty propagation [JCGM 101:2008] to a ratio where a Gaussian random variable appears in the denominator. Since a Gaussian random variable conceivably can take values arbitrarily close to zero, the ratio can become arbitrarily large (in absolute value). If the denominator must be positive, and should have negligible probability of being close to zero, and its relative uncertainty is small (say, less than 10 %), then it is best modeled using a lognormal distribution instead.

Other cases where the lognormal distribution may be a useful model for uncertainty include: (i) situations where standard uncertainties are proportional to measured values; (ii) when replicated determinations of a positive quantity tend to be markedly skewed to the right, or span several orders of magnitude.

5.9 SKEW-NORMAL DISTRIBUTION. The skew-normal distribution generalizes the Gaussian distribution by entertaining a modicum of asymmetry. It is used in Example 8.8.A to model and propagate uncertainties expressed asymmetrically, and the *NIST Uncertainty Machine* uses it to model uncertainties expressed this way. Other asymmetric distributions may be used for the same purpose (for example, the Weibull and the generalized extreme value distributions). Possolo et al. [2019] illustrate the use of the skew-normal distribution in examples from different areas of measurement science.

5.10 UNIFORM (RECTANGULAR) DISTRIBUTION. If a quantity has a uniform (or rectangular) distribution concentrated on an interval of finite length, then the probability of it taking a value in a specified sub-interval of that interval is proportional to the length of this sub-interval. The uniform distribution can be used to model the uncertainty surrounding the atomic weights [Possolo et al., 2018b], in particular for those elements whose standard atomic weights are given explicitly as intervals, as in Example 8.4.A.

5.11 BETA DISTRIBUTION. The beta distribution is mentioned in Example 7.3.C, where it is used to describe the uncertainty associated with the purity of benzoic acid. It is a flexible model for the uncertainty of quantities like mass or amount fractions, which must lie between 0 and 1. Also for other quantities whose values, suitably rescaled and shifted, can be mapped onto the interval (0, 1). The beta probability densities can take many different shapes: bell-shaped (symmetric or asymmetric, skewed to the left or to the right), U, L, or J-shaped. The uniform distribution is a member of this family. The *NIST Uncertainty Machine* offers the (rescaled and shifted) beta distribution as a possible model for input quantities, specified by the mean, standard deviation, and left and right endpoints of its range.

5.12 LINEAR COMBINATIONS OF UNCORRELATED RANDOM VARIABLES. If $\alpha_1, \dots, \alpha_n$ are constants (positive or negative), and X_1, \dots, X_n are uncorrelated random variables with means ξ_1, \dots, ξ_n , and standard deviations $\sigma_1, \dots, \sigma_n$, then the mean value of $Y = \alpha_1 X_1 + \dots + \alpha_n X_n$ is $\alpha_1 \xi_1 + \dots + \alpha_n \xi_n$, and its standard deviation is $(\alpha_1^2 \sigma_1^2 + \dots + \alpha_n^2 \sigma_n^2)^{1/2}$. These facts should help understand the calculations in Example 8.4.A. Note that standard deviations generally should not be added or subtracted. In some cases, it may be meaningful to compute geometric averages of standard deviations, for example $\sqrt{\sigma_1 \sigma_2}$, which is the same as averaging their logarithms and exponentiating the result: $\exp((\log \sigma_1 + \log \sigma_2)/2)$.

5.13 WELCH-SATTERTHWAITE. The measurement model considered in the GUM expresses an output quantity as a function of several input quantities, $Y = f(X_1, \dots, X_n)$, and models all of them as random variables whose probability distributions describe their associated uncertainties.

To produce an approximation to $u_c(y)$, the GUM approximates f by a linear function in a neighborhood of the measured values of the input quantities, $Y \approx \alpha_0 + \alpha_1 X_1 + \dots + \alpha_n X_n$, whence it follows (when the inputs are uncorrelated) that $u_c^2(y) \approx \alpha_1^2 u^2(x_1) + \dots + \alpha_n^2 u^2(x_n)$.

Now, suppose that each $u(x_j) = s_j/\sqrt{m_j}$ is the result of a Type A evaluation, and that x_j is an average of m_j observations whose sample standard deviation is s_j , the true standard deviation being σ_j . In these circumstances, $v_j(s_j/\sigma_j)^2$ is like an outcome of a random variable with a chi-squared probability distribution with $v_j = m_j - 1$ degrees of freedom.

Therefore, the approximate value of $u_c^2(y)$ above is like an outcome of a linear combination of independent chi-squared random variables. There is no closed-form distribution for such linear combination, but it can be approximated by a chi-squared distribution with number of degrees of freedom ν given by the so-called Welch-Satterthwaite formula [Satterthwaite, 1946; Welch, 1947]:

$$\nu = \frac{\left(\sum_{j=1}^n \alpha_j^2 u^2(x_j)\right)^2}{\sum_{j=1}^n \alpha_j^4 u^4(x_j)/v_j} = \frac{u_c^4(y)}{\sum_{j=1}^n \alpha_j^4 u^4(x_j)/v_j}.$$

5.14 WELCH-SATTERTHWAITE — APPLICATION. A determination of nitromethane based on the kinetics of its decomposition in the presence of excess base involved a preliminary experiment to determine the rate constant, which produced the estimate $k = 1.62 \text{ s}^{-1}$ with $u(k) = 0.06 \text{ s}^{-1}$ on 3 degrees of freedom [Skoog et al., 2014, Table 30-2].

The reaction involving the solution to be analyzed was replicated five times, and allowed to take its course over a period of average duration $t = 2 \text{ s}$, with standard uncertainty $u(t) = 0.1 \text{ s}$ on 4 degrees of freedom. The average amount concentration of nitromethane after that time was $c_t = 6.37 \times 10^{-4} \text{ mol/L}$, with standard uncertainty $u(c_t) = 0.08 \times 10^{-4} \text{ mol/L}$, also on 4 degrees of freedom.

The estimate of the initial amount concentration of nitromethane is $c_0 = c_t \exp(kt) = 6.37 \times 10^{-4} \text{ mol/L} \times \exp(1.62 \times 2) = 1.63 \times 10^{-2} \text{ mol/L}$. The sensitivity coefficients are $\partial c_0/\partial c_t = 25.53$ for the final amount concentration, $\partial c_0/\partial k = 0.03253 \text{ mol L}^{-1} \text{ s}$ for the rate constant, and $\partial c_0/\partial t = 0.02635 \text{ mol L}^{-1} \text{ s}^{-1}$ for the reaction time. Therefore, $u_c^2(c_0)/(\text{mol/L})^2 = (25.53 \times 0.08 \times 10^{-4})^2 + (0.03253 \times 0.06)^2 + (0.02635 \times 0.1)^2 = (0.003285)^2$.

The effective number of degrees of freedom supporting $u_c(c_0)$, computed using the Welch-Satterthwaite formula, is

$$\nu = \frac{(0.003285)^4}{(25.53 \times 0.08 \times 10^{-4})^4/3 + (0.03253 \times 0.06)^4/4 + (0.02635 \times 0.1)^2/4} = 7.43.$$

Therefore, a 95 % coverage interval for the true, initial amount concentration of nitromethane, based on the Student's t_v distribution is $1.63 \times 10^{-2} \text{ mol/L} \pm (2.337 \times 0.003\,285 \text{ mol/L})$: that is, ranging from $0.862 \times 10^{-2} \text{ mol/L}$ to $2.40 \times 10^{-2} \text{ mol/L}$.

6 Evaluating Measurement Uncertainty

6.1 Measurement uncertainty is evaluated consistently with a measurement model that relates the measurand to observable values of properties that determine the value of the measurand. For example, in a coulometric measurement, the amount concentration of measurand (expressed in moles per liter, say) is related to measurements of electrical current (passing through a solution containing the measurand) and of time (for a particular chemical reaction to complete).

6.2 The measurement model may be deterministic, as in the direct spectrophotometric measurement of the concentration of bilirubin in serum using the caffeine reagent [Vink et al., 1986] with *measurement equation* $c_{\text{SAMPLE}} = c_{\text{STD}}(A_{465} - A_{528})_{\text{SAMPLE}} / (A_{465} - A_{528})_{\text{STD}}$, where c_{SAMPLE} and c_{STD} denote amount-of-substance concentrations of bilirubin in the sample under measurement and in the standard (for example, NIST SRM 916a Bilirubin), and A_{465} and A_{528} denote the values of absorbance at 465 nm and at 528 nm [Vink et al., 1988].

Or the measurement model may be statistical (also called *observation equation*) as in the measurement of the amount fraction of methane in NIST SRM 1720 Northern Continental Air via gas chromatography with flame-ionization detection (GC-FID), with an analysis function built using errors-in-variables regression [Rhoderick et al., 2016].

6.3 Measurement uncertainty may be evaluated in a *bottom-up* fashion: that is, by identifying all sources of uncertainty that make non-negligible contributions to the uncertainty of the result, and by quantifying the contributions that they make to the overall uncertainty. The results of this process are usually gathered in an *uncertainty budget*. The contributions from these sources are then propagated to the estimate of the measurand, typically through a measurement equation, as explained in the GUM, and as implemented in the *NIST Uncertainty Machine* [Lafarge and Possolo, 2015].

EXAMPLE 6.3.A The gravimetric determination of the mass fraction of mercury in NIST SRM 1641e Mercury in Water included a bottom-up approach to uncertainty evaluation. The measurement equation is

$$w_{1641e} = \frac{m_{3133} w_{3133} m_{\text{spike}}}{m_{\text{spiking soln}} m_{1641e}},$$

where w_{1641e} is the mass fraction of mercury in the material, w_{3133} is the mass fraction of mercury in NIST SRM 3133 Mercury (Hg) Standard solution, m_{3133} is the mass of SRM 3133 used in the first dilution to prepare the intermediate spiking solution, $m_{\text{spiking soln}}$ is the total mass of that solution, m_{spike} is the mass of the intermediate spiking solution used in the second (and final) dilution step to prepare SRM 1641e, and m_{1641e} is the total mass of this solution.

The following uncertainty budget is adapted from Butler and Molloy [2014, Table 1], where “DF” denotes the number of degrees of freedom that each standard uncertainty is based on, and “MODEL” specifies the corresponding probability model for use in the *NIST Uncertainty Machine*.

INPUT	ESTIMATE	STD. UNC.	DF	MODEL
m_{3133}	1.0024 g	0.0008 g	∞	Gaussian
w_{3133}	9.954×10^6 ng/g	0.024×10^6 ng/g	12.5	Student t
$m_{\text{spiking soln}}$	51.0541 g	0.0008 g	∞	Gaussian
m_{spike}	26.0290 g	0.0008 g	∞	Gaussian
m_{1641e}	50 050.6 g	0.1 g	∞	Gaussian

6.4 The contributions from the sources of uncertainty listed in an uncertainty budget may be quantified using statistical methods (usually based on the dispersion of replicated observations of the value of the corresponding quantity), a so-called *Type A evaluation*, or they may be quantified using other methods, a so-called *Type B evaluation*, which may be based on historical information, expert knowledge about measurement methods, etc.

Note that the uncertainty budget in Example 6.3.A does not specify how the contributions from the individual sources of uncertainty were evaluated (Type A or Type B) because they are all treated alike in practice, in accordance with the GUM (4.3.3 and E.3).

6.5 *Top-down* uncertainty evaluations, including interlaboratory studies and comparisons with a reference, provide evaluations of measurement uncertainty without requiring or relying on prior identification and characterization of the contributing sources of uncertainty.

6.6 The evaluation of measurement uncertainty may involve both Type A and Type B evaluations, as well as bottom-up and top-down procedures, in tandem.

EXAMPLE 6.6.A The certified value of the mass fraction of arsenous acid (H_3AsO_3) in NIST SRM 3037 Arsenous Acid (AsIII) Standard Solution is 1.0442 mg/g. The associated uncertainty includes the results of Type A and Type B evaluations of different sources of uncertainty that were combined to produce the reported uncertainty in the form of an expanded uncertainty for 95 % coverage, of 0.0020 mg/g.

The Type A evaluation involved application of the DerSimonian-Laird procedure for consensus building. This procedure was used to blend replicated determinations made for 10 ampoules using the same measurement method, and to gauge the contribution from possible heterogeneity. The Type A evaluation was top-down because it was based on the observed dispersion of the determinations made in the different ampoules, with no attempt to diagnose its root causes.

The Type B evaluation focused on the sources of uncertainty affecting the gravimetric titrimetry (including for the mass of the titrant, among others), and the results of this bottom-up evaluation were summarized in an uncertainty budget.

6.7 *Nominal* and *ordinal* properties are kinds of *categorical* properties, which are qualitative [Agresti, 2019]. NIST SRM 3246 *Ginkgo biloba* (Leaves) assigns the species value *Ginkgo biloba* to the material, based on identifying sequences of nucleobases in its DNA, and expresses the associated uncertainty qualitatively in an ordinal scale that represents the strength of the belief in the assigned value.

FASTA files [Lipman and Pearson, 1985] specify sequences of nucleobases in DNA or RNA, for example for the plant *Ginkgo biloba* in NIST SRM 3246, or sequences of amino acids in proteins, for example for albumin in NIST SRM 972e Bovine Serum Albumin (7 % Solution). These files are plain text files comprising one or more comment lines, followed by lines with strings of standard single letter codes representing nucleobases or amino acids.

FASTQ files [Cock et al., 2009] list nucleobase assigned values and qualify each value assignment with a quality score that can be translated into a probability of incorrect assignment. Example 6.7.C uses FASTQ-style quality scores extracted from a Solexa prb file distributed as part of the Bioconductor package ShortRead [Morgan et al., 2009], to illustrate how the corresponding uncertainties are expressed using locus-specific, discrete probability distributions.

EXAMPLE 6.7.A In NIST SRM 3246 *Ginkgo biloba* (Leaves) the certified species identification is *Ginkgo biloba*, based on sequences of nucleobases from the *psbA-trnH* intergenic spacer and *trnL* intron regions of the plant's DNA. The uncertainty associated with the identity assigned to each nucleobase is expressed in terms of its opposite, *confidence level*, drawn from the ordinal values: *ambiguous, confident, very confident, most confident*.

EXAMPLE 6.7.B NIST SRM 927e Bovine Serum Albumin (7 % Solution) delivers the certified value of the mass concentration of bovine serum albumin (BSA), 67.8 g/L \pm 1.8 g/L. The certified value was derived from the mass concentrations measured for five amino acids (phenylalanine, proline, isoleucine, leucine, and valine), together with the amino acid sequence of mature BSA (residues 25-607, further interlinked with seventeen disulfide bonds). The mass concentrations of those amino acids were determined via amino acid analysis and isotope dilution liquid chromatography/tandem mass spectrometry (ID-LC/MS/MS).

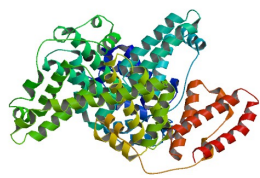


Image from RCSB PDB ([rcsb.org](https://www.rcsb.org)) of PDB ID 3V03

The defining string for the protein comprises 583 letters designating amino acids: DTHKSE...STQTALA. In this case the sequence is not used for identification but for computing both the mass concentration of BSA based on the mass concentrations of the five amino acids mentioned above, and the “theoretical” relative molecular mass of the protein.

EXAMPLE 6.7.C DNA sequencing is the determination of the identity of the nucleobases along a strand of DNA. Each locus along the strand is occupied by one of four nucleobases: adenine (A), guanine (G), cytosine (C), or thymine (T). Genes are sequences of nucleobases that encode the information that biological cells use to manufacture their functional and structural components. The ability to sequence and compare DNA fragments is key to understanding biological function, identifying organisms, and characterizing relationships between them.

The sequencing method developed by Frederick Sanger and collaborators in the late 70s remained the most widely used method during the subsequent forty years

[Sanger and Coulson, 1975; Sanger et al., 1977], and is still in use today, alongside higher-throughput alternatives. Sanger sequencing involves measurement of laser-induced fluorescence of DNA fragments labeled with fluorochromes, and computer processing of the observations.

Some computer programs used for this purpose, for example *Phred* [Ewing et al., 1998; Ewing and Green, 1998], produce not only the identity of the nucleobase at each particular locus along the sequence, but produce also quality scores that can be converted into probabilities expressing the confidence in the identification.

The sequence TTTT**T**AATTGGTTAATCATT**T**TTTTTTTTTAATTTT is one of the examples provided in the Bioconductor package `ShortRead` [Morgan et al., 2009], and it is qualified with quality scores. Thymine (T) was assigned to the seventh location in this sequence (highlighted in boldface above) because thymine had the highest probability given the fluorescence spectrum observed at this location.

The probabilities of the four possible values for the identity of the nucleobase at this location were $\Pr(A) = 0.04$, $\Pr(C) = 0.09$, $\Pr(G) = 0.06$, and $\Pr(T) = 0.81$: they add up to 1 and indeed are a probability distribution over the set $\{A, C, G, T\}$. The entropy of this probability distribution, $-\Pr(A) \log \Pr(A) - \Pr(C) \log \Pr(C) - \Pr(G) \log \Pr(G) - \Pr(T) \log \Pr(T) = 0.69$, is a quantitative evaluation of the uncertainty associated with the identification of the nucleobase at this location: the larger the entropy, the greater the uncertainty.

The entropy of a probability distribution describes the dispersion of the unit of probability, and it is meaningful for probability distributions quantifying the uncertainty associated with quantitative or qualitative measurands. However, in those cases where both the entropy and the standard deviation can be computed, they are not directly comparable because they are expressed in different scales. For example, the entropy of a Gaussian distribution with standard deviation σ is equal to the natural logarithm of σ plus a constant.

7 Expressing Measurement Uncertainty

7.1 The most thorough description of measurement uncertainty is in the form of a probability distribution over the set of possible values of the measurand. However, this is not how measurement uncertainty is usually expressed because (i) the concept of *probability distribution* is abstract, mathematical, and generally unfamiliar, and (ii) the purpose that the measurement result is intended to serve typically does not require an expression of measurement uncertainty of such level of sophistication. When a probability distribution is required for uncertainty propagation, a reasonable choice consistent with the reported uncertainty usually suffices, as illustrated in 8.

7.2 The most common expression of measurement uncertainty for scalar measurands in NIST certificates is an interval believed to include the true value of the measurand with approximately 95 % probability. The meaning of the interval is the same regardless of whether it is of the form $x \pm k_p u_c(x)$, where x denotes the measured value and $u_c(x)$ denotes the associated standard uncertainty, or whether it may have been obtained in some other way, or is not even centered at (and symmetric relative to) the measured value. The aspect of the interval (symmetric or not) and how the interval may have been derived do matter for how the uncertainty will be propagated, as will be discussed in 8.

EXAMPLE 7.2.A The certified mass concentration of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level) is reported as 39.4 ng/mL \pm 0.8 ng/mL with 95 % confidence, signifying that the corresponding true value is believed to lie between 38.6 ng/mL and 40.2 ng/mL with 95 % probability. This is an expression of the state of knowledge that NIST has about this material. It *does not* mean that 95 % of the values measured by users of the material are expected to land inside this interval.

Suppose that users of this SRM make determinations of the mass concentration of that compound in the material and obtain 34.85 ng/mL \pm 2.45 ng/mL as measurement result, where 2.45 ng/mL is an expanded uncertainty for 95 % confidence.

Since this interval and the interval above do not overlap, the users are entitled to conclude that their result differs significantly from NIST's. A formal z -test of the statistical significance of the difference between the user measured and NIST certified values yields a p -value of 0.0004 (refer to A.2 for the corresponding calculations).

The p -value is the probability of observing a difference as large or larger (in absolute value) than the difference that was observed, by chance alone, owing to the vagaries of sampling and measuring the material, if the corresponding true values (NIST's and the users') were identical. A small p -value suggests a significant difference.

The conventional, statistical z -test of hypothesis is based on the difference between the two measured values, divided by the standard uncertainty of the difference. The validity of this test is contingent on the two values being compared conceivably being outcomes of Gaussian random variables, and on their associated standard uncertainties being based on infinitely many degrees of freedom.

Now suppose that another user's measured value is 37.45 ng/mL with expanded uncertainty 1.75 ng/mL, also for 95 % coverage. Since the interval 37.45 ng/mL

± 1.75 ng/mL and NIST's interval overlap, and indeed overlap considerably, the inference may be drawn that this user's result is consistent with NIST's, in the sense that there is no significant difference between the true mean of the user's measurement and the true value according to NIST.

This is not necessarily the case because such inference is based on the all-too-common misinterpretation of the overlap of the coverage intervals as being the same as the overlap of the probability distributions that they summarize.

The combined standard uncertainty for NIST is $(0.8 \text{ ng/mL})/2 = 0.4 \text{ ng/mL}$ because the certificate states that the coverage factor is 2. Assuming that the same coverage factor applies to the user, the corresponding standard uncertainty is $1.75 \text{ ng/mL}/2 = 0.875 \text{ ng/mL}$.

The difference between the two measured values is $39.4 \text{ ng/mL} - 37.45 \text{ ng/mL} = 1.95 \text{ ng/mL}$. The standard uncertainty of the difference between the measured values is the square root of the sum of the individual, squared standard uncertainties, $(0.4^2 + 0.875^2)^{1/2} \text{ ng/mL} = 0.962 \text{ ng/mL}$.

The test statistic is the standardized difference, $1.95/0.962 = 2.03$. The p -value of the test is the probability of a Gaussian random variable with mean 0 and standard deviation 1 being either smaller than -2.03 or larger than $+2.03$. This probability is 4 %, which most people would regard as suggesting a significant difference. [A.2](#) lists R code implementing this computation.

EXAMPLE 7.2.B The certified measurement result for ascorbic acid (vitamin C) in NIST SRM 3280 Multivitamin/Multielement Tablets is the interval $42.2 \text{ mg/g} \pm 3.7 \text{ mg/g}$, which is believed to include the corresponding true value with 95 % probability. The certificate explains that 3.7 mg/g is the expanded uncertainty computed as 3.15 times the standard uncertainty associated with this mass fraction.

Since the coverage factor is $k_{95\%} = 3.15$ (rather than 2) suggests that the underlying probability distribution that fully describes the uncertainty associated with this mass fraction is a Student's t distribution with approximately 3 degrees of freedom, which is the output of `kInverter(k=3.15, coverage=0.95)`. R function `kInverter`, which is defined in [A.3](#), computes the number of degrees of freedom that correspond to a specified coverage factor k_p and coverage probability p .

EXAMPLE 7.2.C NIST SRM 3161a Tin (Sn) Standard Solution (Lot No. 140917) is a solution with $10.011 \text{ mg/g} \pm 0.025 \text{ mg/g}$ of tin that is part of a collection of (more than 60) similar single element solutions intended to serve as calibrants for spectrometry. The assigned value results from blending gravimetric and inductively coupled plasma optical emission spectrometric (ICP-OES) determinations [[Possolo et al., 2018a](#)].

The associated expanded uncertainty for 95 % coverage, 0.025 mg/g , comprises contributions from sources of uncertainty that are specific to each of those two measurement methods, and also a contribution, τ , or *dark uncertainty* [[Thompson and Ellison, 2011](#)], expressing differences between methods, that is evaluated in a top-down fashion. [A.4](#) lists R code to compute an estimate of τ in this case.

The relative expanded uncertainties of SRMs in this class range from 0.14 % to 0.38 % approximately, with τ often being the predominant component. In fact, in

more than half of the NIST single-element spectrometric solutions, τ is at least twice as large as the contributions that gravimetry and ICP-OES make individually. Users will be unable to perceive dark uncertainty if they should measure NIST SRM 3161a under repeatability conditions, because dark uncertainty manifests itself only when independent measurements of the same measurand are compared that have been made using different methods, or using different instruments, possibly in different laboratories. (“Under repeatability conditions” means that the same measurement procedure was used by the same technician, using the same instrument under the same operating conditions at the same location, and over a short period of time.)

EXAMPLE 7.2.D Li et al. [2014] used NIST SRM 927e Bovine Serum Albumin (7 % Solution), which is introduced in Example 6.7.B, to validate a new analytical method to determine the mass concentration of proteins in solution without the need for specific protein calibration.

The validating experiments, conducted on five different days, produced these estimates of the mass concentration of the bovine serum albumin in NIST SRM 927e: 66.3 g/L, 67.6 g/L, 63.6 g/L, 66.5 g/L, 64.9 g/L. Their average was reported as 65.8 g/L, and the reported uncertainty was 1.6 g/L [Li et al., 2014, Figure 5].

Li et al. [2014] do not explain the meaning of the reported uncertainty. Since the standard deviation of those five determinations is 1.6 g/L, the reported uncertainty may be this standard deviation. If it should be so, then the reported uncertainty is not directly comparable with the uncertainty listed in the certificate of SRM 927e, 1.38 g/L, because this is an expanded uncertainty for 95 % coverage with coverage factor $k = 2$.

To evaluate the agreement between the determinations made using the method that Li et al. [2014] propose and the certified value, a t -test may be employed, with the following criterion (where an additional significant digit has been carried both for the average of the five determinations and for their standard deviation, and where the measurement units cancel):

$$t = \frac{65.78 - 67.38}{\sqrt{\left(\frac{1.55}{\sqrt{5}}\right)^2 + \left(\frac{1.38}{2}\right)^2}} = -1.64.$$

The first term under the square root sign in the denominator, $(1.55/\sqrt{5})^2$, is the squared standard uncertainty of the average of the five determinations. The second term is the squared standard uncertainty associated with the certified value (the “2” in $1.38/2$ is the coverage factor).

The reference probability distribution against which the value of the test criterion, $t = -1.64$, should be compared, is a Student’s t distribution with ν degrees of freedom given approximately by the Welch-Satterthwaite formula (5.13):

$$v = \frac{\left[\left(\frac{1.55}{\sqrt{5}} \right)^2 + \left(\frac{1.38}{2} \right)^2 \right]^2}{\frac{\left(\frac{1.55}{\sqrt{5}} \right)^4}{4} + \frac{\left(\frac{1.38}{2} \right)^4}{60.4}} = 15.7.$$

The “4” in the denominator is the number of degrees of freedom of the standard deviation of the replicated determinations, and the “60.4” is the number of degrees of freedom corresponding to $k = 2$, as produced by `kInverter(k=2, coverage=0.95)` (A.3) when one wishes to interpret the coverage factor $k = 2$ as originating in a Student’s t distribution rather than approximately as originating in a Gaussian distribution

The p -value of the test is 0.12 (probability of drawing a value with absolute value larger than 1.64 from a Student’s t distribution with 15.7 degrees of freedom), which suggests that there is no significant difference between the results of the method that Li et al. [2014] propose and the certified value.

7.3 When the probability distribution used to describe measurement uncertainty is asymmetrical, as in Examples 7.3.A, 7.3.B, 8.8.A, and 8.10.A, it is generally possible to produce an expanded uncertainty $U_p(x)$ such that $x \pm U_p(x)$ has the same meaning as when the distribution is symmetrical, and it is also possible to compute a meaningful standard uncertainty $u_c(x)$. In such cases, however, the implied coverage factor $U_p(x)/u_c(x)$ is neither particularly meaningful nor useful.

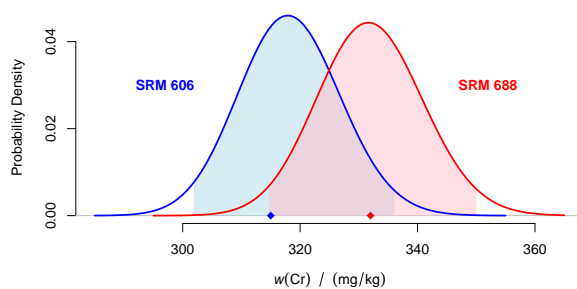
It should also be noted that when the probability distribution used to describe measurement uncertainty is asymmetrical, the shortest coverage interval with specified coverage probability is not centered at the measured value. The *NIST Uncertainty Machine* can model and propagate uncertainties expressed asymmetrically.

EXAMPLE 7.3.A The certified mass fraction of chromium in NIST SRM 606 Trace Elements in Basalt Glass is 315 mg/kg, with associated standard uncertainty 9 mg/kg, and a 95 % coverage interval that ranges from 302 mg/kg to 336 mg/kg. The interval is markedly asymmetrical, with the sub-interval to the right of the measured value being 1.6 times longer than the sub-interval to the left of the measured value. SRM 606 is a glass obtained after several cycles of melting and stirring the molten basalt of NIST SRM 688 Basalt Rock, whose certificate (issued 37 years prior to the certificate of SRM 606) puts the mass fraction of chromium at 332 mg/kg, with uncertainty 9 mg/kg. Since the certificate states that this uncertainty is “based on judgment,” one may conclude that it is the result of a Type B evaluation.

The example serves to illustrate how two measurement results may be compared when the respective uncertainties are expressed using either asymmetric coverage intervals, or asymmetric probability distributions. Next we will show that this can be accomplished very easily. Examples 8.8.A and 8.10.A discuss similar, but more challenging cases.

The simplifying fact is that the uncertainty surrounding the mass fraction of chromium in SRM 606, $w_{\text{SRM606}}(\text{Cr})$, can be described fairly accurately by a log-normal distribution. In other words, that $\log w_{\text{SRM606}}(\text{Cr})$ is like an outcome of

a Gaussian random variable with mean 5.76 and standard deviation 0.0273. By analogy, a lognormal distribution will also be used to describe the uncertainty associated with the mass fraction of chromium in SRM 688, so that $\log w_{\text{SRM688}}(\text{Cr})$ is like an outcome of a Gaussian random variable with mean 5.80 and standard deviation 0.0271.



The light blue and pink regions cover 95 % of the areas under the respective curves. The certified values are represented by small diamonds.

The challenge posed by the asymmetric distributions is resolved very simply by comparing the results on a logarithmic scale, because on this scale the lognormal distributions become Gaussian distributions, and a conventional statistical test of hypothesis applies.

The z-test of equality of the true values of $\log w_{\text{SRM606}}(\text{Cr})$ and of $\log w_{\text{SRM688}}(\text{Cr})$ is based on the standardized difference between the logarithms of the measured values: $(5.76 - 5.80)/(0.0274^2 + 0.0273^2)^{1/2} = -1.03$. The corresponding *p*-value is 0.30.

The *p*-value is the probability of a Gaussian random variable deviating more than 1.03 standard deviations from its mean in either direction, and means that a difference this large or larger (in absolute value) has appreciable probability of occurring due to the vagaries of sampling alone, when there is no actual difference.

The conclusion, therefore, is that the true values are not significantly different. (A z-test conducted with 5 % probability of erroneously rejecting the hypothesis of no difference would suggest rejecting this hypothesis only when the absolute value of the standardized difference exceeds 1.96.)

The only aspect remaining to be clarified is how these lognormal distributions were fitted to the measurement results. A.6 lists R code implementing the computations described next.

A lognormal distribution is determined by the values of two parameters, μ and $\sigma > 0$. Since the certificate of SRM 688 lists the measured value $w_{\text{SRM688}}(\text{Cr})$ and the standard uncertainty $u(w_{\text{SRM688}}(\text{Cr}))$, by setting these equal to the mean and standard deviation of a lognormal distribution, its parameters can be reproduced exactly, by solving these two simultaneous, nonlinear equations: $\exp(\mu + \sigma^2/2) = w_{\text{SRM688}}(\text{Cr})$ and $(\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2) = u(w_{\text{SRM688}}(\text{Cr}))$. The solution is $\mu = \log(\eta/((\tau/\eta)^2 + 1)^{1/2})$, and $\sigma = (\log((\tau/\eta)^2 + 1))^{1/2}$, where $\eta = w_{\text{SRM688}}(\text{Cr})$ and $\tau = u(w_{\text{SRM688}}(\text{Cr}))$.

The situation is different for SRM 606, because its certificate provides four pieces of information for the mass fraction of chromium: measured value, standard uncertainty, and the endpoints of a 95 % coverage interval. These lead to an overdetermined system of nonlinear equations for μ and σ that can be solved by weighted least squares, for example.

It is advisable to give the greater weights to the endpoints of the coverage interval, best to characterize the tails of the distribution, which are its most influential feature for the purpose of uncertainty propagation. The details of these calculations are provided in A.6.

If one would have chosen to ignore the information provided by the endpoints of the asymmetrical coverage interval listed for NIST SRM 606, then the comparison could have been done more simply as follows, considering that the standard uncertainty of the natural logarithm of a measured value is approximately equal to its relative standard uncertainty:

$$z = \frac{(\log(w.606) - \log(w.688))}{\sqrt{((uw.606/w.606)^2 + (uw.688/w.688)^2)}} \\ 2 * \text{pnorm}(-\text{abs}(z))$$

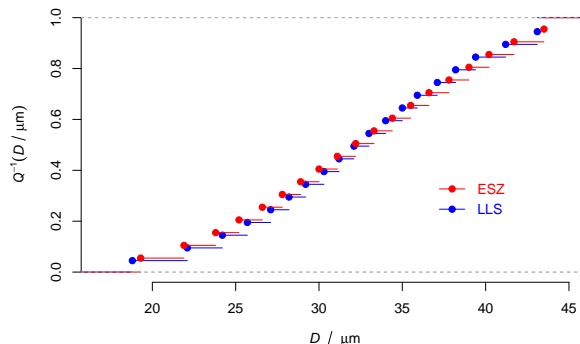
This simplified approach yields p -value 0.18, while the more sophisticated approach described previously had yielded p -value 0.30. The conclusion, of no significant difference, would be the same in either case.

EXAMPLE 7.3.B NIST SRM 2780a (Hard Rock Mine Waste) is a powdered version of the history of mining in central Colorado, comprising materials collected from waste piles of abandoned mines there. The assigned values are mass fractions of many elements, and the associated uncertainties are expressed as uncertainty intervals believed to include the true values they pertain to, with 95 % confidence. The certified value for chromium is 205 mg/kg, with the uncertainty interval ranging from 194 mg/kg to 233 mg/kg. The interval is not centered at the certified value, but is asymmetric relative to the certified value, with the “right” expanded uncertainty appreciably larger than its “left” counterpart.

EXAMPLE 7.3.C NIST PS1 Primary Standard for quantitative NMR (Benzoic Acid) is intended for use as a primary standard for quantitative nuclear magnetic resonance (qNMR) methods [Nelson et al., 2018]. Differently from most other NIST reference materials, this one is available only under a Material Transfer Agreement. The certified value is 0.999 92 g/g, and the asymmetrical interval ranging from 0.999 86 g/g to 0.999 96 g/g is believed to include the true value of the purity with 95 % probability. The uncertainty surrounding the certified value can also be characterized approximately by means of an asymmetrical beta distribution with values of the shape parameters specified in the respective certificate.

7.4 The uncertainty surrounding estimates of measurands that are functions may be expressed either (i) for particular values of the function, or (ii) in the form of a confidence envelope surrounding the measured function as a whole over the range of values of interest. The first option was used for absorbance spectra in NIST SRM 2034 Holmium Oxide Solution Wavelength Standard (240 nm to 650 nm), at 14 wavelengths, and for particle size distributions in NIST SRM 1003c Glass Beads (Particle Size Distribution), at 19 percentiles. The second option was used for Raman spectra in NIST SRM 2242a Relative Intensity Correction Standard for Raman Spectroscopy (532 nm Excitation).

EXAMPLE 7.4.A The measurand in NIST SRM 1003c Glass Beads (Particle Size Distribution) is a function Q such that if $0 < p < 1$, then $Q(p)$ is a length corresponding to the proportion p of the beads in the material (soda-lime glass beads) of diameter less than or equal to $Q(p)$. The function has been certified at each of nineteen particular values of its argument: for example, $Q(0.50) = 32.1 \mu\text{m}$ (hence this is the median diameter of the beads), with $u(Q(0.50)) = 0.50 \mu\text{m}$ and $U_{95\%}(Q(0.50)) = 1.0 \mu\text{m}$.



Let us examine this figure in some detail, because it conveys a lot of information. The figure depicts two functions, one in blue, the other in red. By rights, each pair of vertically nearby blue and red little horizontal line segments and dots should be at the same height: they are shifted slightly from one another — one upwards, the other downwards — only to make it easy to perceive the differences between them. Consider the blue function, which summarizes the results of measuring the particles using a Laser Light Scattering (LLS) instrument. Essentially the same explanations apply to the red function, which summarizes the results of measuring the particles using an Electric Sensing Zone (ESZ) instrument.

The blue function is represented by dots and line segments, each of which extends to the right of the corresponding dot. The function remains constant as D increases over the range corresponding to a line segment, and then it jumps to the blue dot immediately above where the line segment ends.

The jumps are a consequence of how the measured values were recorded. The certificate does not list the diameters of individual particles. It says, for example, that 15 % of the particles have diameter $24.2 \mu\text{m}$ or smaller, and that 20 % of the particles have diameter $25.7 \mu\text{m}$ or smaller.

Therefore, at $25.7 \mu\text{m}$ (on the horizontal axis), the function jumps from 0.15 to 0.20. Before it jumps, that is, for values of D between $24.2 \mu\text{m}$ and $25.7 \mu\text{m}$, it stays constant at 0.15.

The vertical axis of this plot could just as well have been labeled $F(D)$, where F denotes a cumulative probability distribution function. For example, $F(32.1) = 0.5$, which means that 50 % of the particles have diameter $32.1 \mu\text{m}$ or smaller.

In other words, F and Q are two sides of the same coin, two different ways of describing the same information. Using mathematical language, we can say that F and Q are the mathematical inverses of one another: $F(Q(p)) = p$ for every $0 < p < 1$, and $Q(F(D)) = D$ for every value D of the diameter.

Besides the certified values and associated uncertainties, the certificate lists two sets of values of Q for each of ten bottles, one set determined using LLS, the other

using ESZ. The figure above shows the results for Bottle 1 as listed in Tables 3 and 4 of the certificate. The largest vertical distance between the blue and red functions is 0.05. Even though the certificate does not state the number of particles whose diameters were measured, based on other information we believe that it was 1000.

The information just stated suffices to carry out the so-called Kolmogorov-Smirnov two-sample statistical test [Conover, 1999] of whether the true functions corresponding to the observed blue and red functions are identical. A.7 lists R code to perform this test, which yielded a p -value of 0.013.

This means that, if the true functions indeed were identical, then the probability of seeing a difference this large or larger, owing to the vagaries of sampling the bottle, is 0.013. Since this is a small probability, we conclude that the difference is statistically significant. However, the difference is very small anyway, and it is inconsequential for the material's intended purpose.

EXAMPLE 7.4.B NIST SRM 2657a Oxygen in Nitrogen is certified for the amount fraction of oxygen in nitrogen. The certified value is 1.9650 cmol/mol, with expanded uncertainty 0.0055 cmol/mol for 95 % coverage. The units of this material are gas mixtures in 6 L cylinders. Value assignment, and the corresponding uncertainty evaluation, for this material and for similar gas mixture reference materials, have been done as described by Guenther and Possolo [2011], improving on [ISO 6143:2001].

The instrument used to measure oxygen in gas mixtures without interfering compounds exploits the fact that the oxygen molecule is paramagnetic. To account for any temporal drift of the instrument, the measurement data used for calibration and for value assignment consists of ratios of instrumental indications to an analytical control, designated as the SRM *lot standard*. Each ratio is the value of the instrumental response for the sample under measurement, divided by the corresponding, contemporaneous response for the lot standard.

First, an analysis function is built using NIST certified amount fractions and associated uncertainties for appropriate primary standard gas mixtures (PSMs), and their corresponding ratios of instrumental indications, obtained in the same manner as for the material of interest. Second, this function is applied to the ratio obtained for the mixture in each cylinder, to produce an estimate of its amount fraction of oxygen.

Since the batch was found to be sufficiently homogeneous for the intended purpose, all cylinders were assigned the same value, which was the average of the amount fractions computed for the individual cylinders.

The associated uncertainty was evaluated by application of the parametric statistical bootstrap, which allows computing a 95 % coverage interval centered at the assigned value directly, without involving the standard uncertainty or a coverage factor.

The corresponding expanded uncertainty is half the length of this interval. The standard uncertainty is the standard deviation of the amount fractions computed for the cylinders in the batch, and the implied coverage factor k is computed as the ratio of the expanded and standard uncertainties.

8 Propagating Measurement Uncertainty

8.1 The certificates of some older reference materials produced by NIST provide certified values and evaluations of measurement uncertainty whose precise nature is not specified. Typically, these materials were evaluated before the standards for reporting uncertainty that are in use today had been formulated and adopted. In such cases, it is recommended that the reported uncertainty be treated as if it were an evaluation of standard uncertainty. If the certified value and its uncertainty are to be used in a Monte Carlo method that requires specification of a probability distribution, then this distribution may be assumed to be Gaussian, unless prior knowledge of the measurand in this type of material, or general knowledge of the relevant chemistry or geochemistry, suggests that a different model should be chosen.

EXAMPLE 8.1.A NIST SRM 278 Obsidian Rock is a finely powdered obsidian extruded during the most recent eruption (around 640 A.D.) of the Newberry Volcano, covering about 2.6 km² of the Newberry National Volcanic Monument, near Bend, Oregon. The certificate lists certified values for the mass fraction of lead, $w(\text{Pb}) = 16.4 \text{ mg/kg} \pm 0.2 \text{ mg/kg}$, and of uranium, $w(\text{U}) = 4.58 \text{ mg/kg} \pm 0.04 \text{ mg/kg}$. The certificate does not explain the meaning of these uncertainties, saying only that they are “based on judgment and represent an evaluation of the combined effects of method imprecision, possible systematic errors among methods and material variability.” Suppose that we wish to propagate them to the ratio $w(\text{Pb})/w(\text{U})$, for example in a geochemical calculation, and that we will do so using the *NIST Uncertainty Machine* as follows, where the numerator and denominator of the ratio appear as w_{Pb} and w_{U} :

The screenshot shows the NIST Uncertainty Machine interface with the following settings:

- 1. Select Inputs & Choose Distributions:**
 - Number of input quantities: 2
 - Names of input quantities: w_{Pb} , w_{U}
 - w_{Pb} : Gaussian (Mean, StdDev) with Mean 16.4 and StdDev 0.2
 - w_{U} : Gaussian (Mean, StdDev) with Mean 4.58 and StdDev 0.04
 - Correlations: Off
- 2. Choose Options:**
 - Number of realizations of the output quantity: 1000000
 - Random number generator seed: 33
 - Symmetrical coverage intervals: On
- 3. Write the Definition of Output Quantity:**
 - Definition of output quantity (R expression): $w_{\text{Pb}} / w_{\text{U}}$

The estimates of the ratio produced according to the GUM, and by application of the Monte Carlo method of the GUM-S1 [JCGM 101:2008], are both 3.58, and the evaluations of standard uncertainty, 0.054, also agree.

NIST Technical Note 1297 [Taylor and Kuyatt, 1994] recommends that the corresponding expanded uncertainty be computed as $2 \times 0.054 = 0.11$. This happens to agree with the expanded uncertainty implied by the 95 % coverage interval produced by the Monte Carlo method, (3.48, 3.69).

If the Gaussian models should be replaced by lognormal models, a possibility that the *NIST Uncertainty Machine* also offers, then the results are the same as for the Gaussian models (which happens in this case but need not happen in general). The

lognormal model is the recommended alternative to the Gaussian model when the quantity of interest is positive by definition, appears in the denominator of a fraction, and the relative uncertainty is small — say, less than 10 %, which is the case here, both for lead (1.2 %) and for uranium (0.87 %). Refer to 5.8 for the reason why.

8.2 The simplest instance of uncertainty propagation for a measurement result listed in the certificate of a NIST reference material involves a quantity X whose value x is certified, and it is qualified with an evaluation of expanded uncertainty for 95 % coverage, $U_{95\%}(x)$, whose coverage factor $k_{95\%}$ also is listed. The corresponding standard uncertainty therefore is $u(x) = U_{95\%}(x)/k_{95\%}$. If $k_{95\%} = 2$, then for purposes of uncertainty propagation where a probability model is required, X may be modeled as a Gaussian random variable with mean equal to the measured value and standard deviation equal to the associated standard uncertainty. If $k_{95\%} > 2$, then a Student's t distribution should be used instead.

EXAMPLE 8.2.A The certified value of the mass fraction of chromium in NIST SRM 3112a Chromium (Cr) Standard Solution (Lot No. 030730) is $w = 9.922$ mg/g, with expanded uncertainty $U_{95\%}(w) = 0.025$ mg/g. The certificate states that the coverage factor was $k = 2.20$. The corresponding number of degrees of freedom can be computed using R function `kInverter` defined in A.3, as `kInverter(k=2.20, coverage=0.95)`, which yields 11.

The probability distribution that may be used to model the uncertainty surrounding the true value of that mass fraction is a Student's t distribution with mean 9.922 mg/g, standard deviation $u(w) = (0.025 \text{ mg/g})/2.20 = 0.0114$ mg/g, and 11 degrees of freedom.

Suppose one needs to estimate the true value of $\log_{10}(w)$ and characterize the associated uncertainty, possibly to depict them in a graphical display with a logarithmic scale. Since the logarithm is a nonlinear function, we will rely on the Monte Carlo method implemented in the *NIST Uncertainty Machine*, which draws a large sample from the probability distribution of the transformed mass fraction, by taking the following steps, considering that the measurement model is $y = \log_{10}(w)$, thus with a single input quantity:

- Under 1. *Select Inputs & Choose Distributions*, type w as name for the single input quantity.
- Select *Student t (Mean, StdDev, No. of degrees of freedom)* from the drop-down menu, and enter 9.922, 0.0114, and 11 into the three rectangular boxes that appear to the right of the drop-down menu.
- In the box under 3. *Write the Definition of Output Quantity*, type the right-hand side of the measurement equation, $\log_{10}(w)$.
- Click *Run the computation*.

The resulting sample of 10^6 of values of the output y has average 0.9966, standard deviation 0.0005, and a 95 % coverage interval for the true value of $\log_{10}(w/(mg/g))$ ranges from 0.9956 to 0.9976.

8.3 NIST may assign non-certified values to properties of reference materials. According to [Beauchamp et al. \[2020\]](#), the documentation accompanying a non-certified value “must not imply that NIST asserts that a non-certified value represents the true value of the measurand.” (Between 1987 and until [Beauchamp et al. \[2020\]](#) came into effect, NIST non-certified values were called either “NIST Reference Values” or “NIST Information Values” [[May et al., 2000](#)].)

While mindful of such limitation, and in particular aware of the fact that the reported uncertainty associated with a non-certified value reflects contributions from only some of the possibly important sources of uncertainty, a user may still wish to propagate the reported uncertainty.

EXAMPLE 8.3.A The reference value of the diameter of silver nanoparticles in NIST RM 8017 (Polyvinylpyrrolidone Coated Silver Nanoparticles), assumed to be spherical, measured using transmission electron microscopy is $D = 74.6$ nm with expanded uncertainty $U_{95\%}(D) = 3.8$ nm, where $U_{95\%}(D) = k_{95\%}u(x)$ with $k_{95\%} = 2$. To compute a 95 % coverage interval for the volume V of these particles, take the following steps:

- Compute $u(D) = (3.8 \text{ nm})/2 = 1.9$ nm;
- Compute the corresponding number of degrees of freedom using the R function defined in [A.3](#): `kInverter(k=2, coverage=0.95)`, which produces $\nu = 60.4$;
- Use the *NIST Uncertainty Machine* to evaluate the uncertainty of $V = (4/3)\pi(D/2)^3$, assigning to the input representing D a Student’s t distribution with mean 74.6 nm, standard deviation 1.9 mg/g, and 60.4 degrees of freedom, and request symmetrical coverage intervals.

The resulting 95 % coverage interval for the true value of the volume ranges from $185 \times 10^3 \text{ nm}^3$ to $250 \times 10^3 \text{ nm}^3$.

The screenshot shows the NIST Uncertainty Machine interface with the following settings:

- 1. Select Inputs & Choose Distributions:**
 - Number of input quantities: 1
 - Names of input quantities: D
 - D: Student t (Mean, StdDev, No. of degrees of freedom) with values 74.6, 1.9, and 60.4.
 - Correlations: Off
- 2. Choose Options:**
 - Number of realizations of the output quantity: 1000000
 - Random number generator seed: 53
 - Symmetrical coverage intervals: On
- 3. Write the Definition of Output Quantity:**
 - Definition of output quantity (R expression): $(4/3)*\pi*(D/2)^3$

8.4 Consider the measurement model $Y = \kappa X_1^{\alpha_1} \dots X_n^{\alpha_n}$ (a multiple of a product of powers of the input quantities), where κ and $\alpha_1, \dots, \alpha_n$ are known constants (positive or negative numbers, not necessarily integers), and Y is the output quantity, whose true value is the object of measurement. The $\{X_i\}$ are the input quantities, for which estimates and uncertainty evaluations are available. Many models arising in practice are

of this form because many quantities are ratios of products of other quantities (hence with the $\{\alpha_i\}$ all being either +1 or -1).

Now suppose that the measurement results for the input quantities comprise measured values x_1, \dots, x_n and associated expanded uncertainties for 95 % coverage, $U_{95\%}(x_1), \dots, U_{95\%}(x_n)$, whose coverage factors $k_{95\%,1}, \dots, k_{95\%,n}$ also are given. Therefore, the corresponding standard uncertainties are $u(x_1) = U_{95\%}(x_1)/k_{95\%,1}, \dots, u(x_n) = U_{95\%}(x_n)/k_{95\%,n}$. If the relative standard uncertainties $\{u(x_i)/x_i\}$ all are small (say, less than 10 %), then the relative standard uncertainty of the output quantity is such that (GUM §5.1.6)

$$\left(\frac{u_c(y)}{y}\right)^2 \approx \left(\alpha_1 \frac{u(x_1)}{x_1}\right)^2 + \dots + \left(\alpha_n \frac{u(x_n)}{x_n}\right)^2.$$

Without making additional assumptions about the input quantities, it is impossible to produce a valid coverage interval for the true value of the output quantity. However, in many cases an interval of the form $y \pm 2u_c(y)$ will have coverage probability 95 % approximately [Freedman et al., 2007, Page 81], including in cases where the probability distribution of the output quantity is markedly skewed (that is, asymmetrical), supporting Taylor and Kuyatt [1994]’s recommendation.

EXAMPLE 8.4.A The certified value of the mass fraction, w , of 25-hydroxyvitamin D₃ in NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level) is 38.6 ng/g, with standard uncertainty $u(w) = 0.4$ ng/g. The corresponding amount-of-substance concentration is $c = w\rho/M$, where w denotes the mass fraction of the measurand in the solution, ρ denotes the volumetric mass density of the solution, and M denotes the molar mass of the measurand, whose magnitude is approximately equal to the magnitude of its relative molecular mass (or molecular weight).

The calculations described next, which produce the standard uncertainty, $u(c)$, associated with the amount-of-substance concentration c , are implemented in R code listed in A.8.

The volumetric mass density of the solution was determined in triplicate using the Lang-Levy pipet method [Sniegoski and Moody, 1979], yielding 1.022 30 g/mL, 1.022 30 g/mL, and 1.022 26 g/mL. The estimate of this input quantity is the average of these determinations, $\rho = 1.022 29$ g/mL. Since the sample standard deviation of the determinations is $2.309 40 \times 10^{-5}$ g/mL, we have the Type A evaluation $u(\rho) = 2.309 40 \times 10^{-5}$ g/mL/ $\sqrt{3} = 0.000 013$ g/mL (GUM 4.2.3).

The molecular formula of the measurand is C₂₇H₄₄O₂. The standard atomic weights of carbon, hydrogen, and oxygen, all are intervals: [12.0096, 12.0116], [1.00784, 1.00811], and [15.99903, 15.99977], respectively (from CIAAW, *Atomic weights of the elements 2019*, www.ciaaw.org).

The atomic weights of these elements may be modeled as random variables uniformly distributed over their respective intervals [Possolo et al., 2018b]. In this conformity, one may use as their estimates the corresponding mean values (which are the mid-points of those intervals), and, because the distributions are assumed to be uniform (or rectangular), the associated standard uncertainties are the length of

the intervals divided by $\sqrt{12}$. Therefore,

$$\begin{aligned} M/(\text{g/mol}) &= 27(12.0096 + 12.0116)/2 + 44(1.00784 + 1.00811)/2 \\ &\quad + 2(15.99903 + 15.99977)/2 = 400.6359, \text{ and} \\ u(M)/(\text{g/mol}) &= \left[(27^2)(12.0116 - 12.0096)^2/12 \right. \\ &\quad + (44^2)(1.00811 - 1.00784)^2/12 \\ &\quad \left. + (2^2)(15.99977 - 15.99903)^2/12 \right]^{1/2} = 0.016. \end{aligned}$$

To elucidate the foregoing calculation of $u(M)$, consider these facts:

- M is modeled as a random variable so that $u(M)$ can be evaluated as M 's standard deviation: that is, $u^2(M)$ is M 's variance.
- The variance of a sum of uncorrelated random variables is equal to the sum of their variances, and M is such a sum. (In general, the standard deviation of a sum of random variables is *not* equal to the sum of their standard deviations).
- If X is a random variable with variance σ^2 and a is a constant, then the variance of aX is $a^2\sigma^2$. This is the reason why the factors 27^2 , 44^2 , and 2^2 appear on the right-hand side above.
- The other terms that appear on the right-hand side, $(12.0116 - 12.0096)^2/12$, $(1.00811 - 1.00784)^2/12$, and $(15.99977 - 15.99903)^2/12$ are the variances (that is, squared standard deviations) of the random variables that represent the atomic weights of carbon, hydrogen, and oxygen.

The estimate of the amount-of-substance concentration is

$$c = \frac{(38.6 \text{ ng/g})(1.022 29 \text{ g/mL})}{400.6359 \text{ g/mol}} = 98.49 \text{ nmol/L},$$

and the associated standard uncertainty is

$$\begin{aligned} u(c) &\approx c \left[\left(\frac{u(w)}{w} \right)^2 + \left(\frac{u(\rho)}{\rho} \right)^2 + \left(\frac{u(M)}{M} \right)^2 \right]^{1/2} \\ &= (98.5 \text{ nmol/L}) \left[\left(\frac{0.4}{38.6} \right)^2 + \left(\frac{0.000013}{1.02229} \right)^2 + \left(\frac{0.016}{400.6359} \right)^2 \right]^{1/2} \\ &= 1.02 \text{ nmol/L}. \end{aligned}$$

To be able to use the *NIST Uncertainty Machine* to the same effect requires an answer to this question: how to model the volumetric mass density ρ ?

By rights, it ought to be modeled as a Student's t distribution with 2 degrees of freedom (which is the number of degrees of freedom of $u(\rho)$), with standard deviation $u(\rho) = 0.000 013 \text{ g/mL}$, and mean $1.022 29 \text{ g/mL}$.

The problem is that a Student's t distribution with 2 degrees of freedom has infinite variance, hence infinite standard deviation, which cannot be reconciled with the finite value for $u(\rho)$ above.

The *NIST Uncertainty Machine* offers two options for specifying a Student's t model for a quantity surrounded by uncertainty. One option involves specifying the mean,

standard deviation, and number of degrees of freedom, which has to be greater than 2. This option is not practicable in this case because $u(\rho)$ is based on only 2 degrees of freedom.

The other option involves specifying the center of symmetry (which is the median), the scale (which controls the spread of values drawn from this distribution), and the number of degrees of freedom (which control the heaviness of the tails). The center and the scale remain meaningful whatever the number of degrees of freedom may be. When the number of degrees of freedom ν is greater than 2, the standard deviation of the distribution is $\sqrt{\nu/(\nu-2)}$ times larger than the scale parameter. When $\nu = 2$ the standard deviation is infinite, and the best one can do is to set the scale parameter equal to $u(\rho)$.

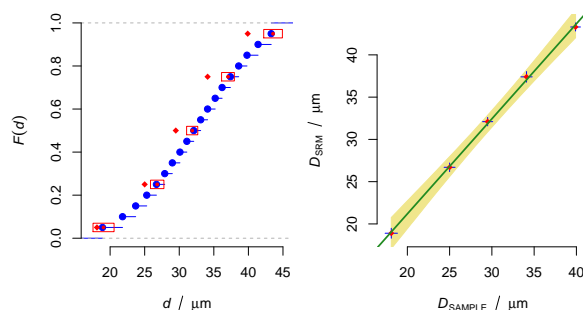
With these choices, the Monte Carlo method as implemented in the *NIST Uncertainty Machine* reproduces the results above and produces a 95 % coverage interval for the true value of c ranging from 96.49 nmol/L to 100.5 nmol/L, thus with coverage factor computed as $k = ((100.5 - 96.49)/1.02)/2 = 1.97$.

8.5 The value assigned to some NIST reference materials is a function, not a scalar quantity or a qualitative property, as already mentioned in 7.4. In Example 7.4.A the measurand is a *quantile* function, the mathematical inverse of a cumulative probability distribution function [DeGroot and Schervish, 2012]. In SRM 1003c Glass Beads (Particle Size Distribution), the uncertainty evaluations were made pointwise, at specific values of the quantile function, with coverage intervals meant to be taken individually, not simultaneously. In Example 8.5.A, the uncertainties reported in the certificate are propagated to the results of a calibration performed using the SRM.

EXAMPLE 8.5.A NIST SRM 1003c Glass Beads (Particle Size Distribution) may be used to calibrate a system used to measure particle size. Suppose that the user determined the 5th, 25th, 50th, 75th, and 95th percentiles of a sample of these beads, and that they turned out to be $18.1 \mu\text{m} \pm 0.73 \mu\text{m}$, $25.0 \mu\text{m} \pm 0.65 \mu\text{m}$, $29.5 \mu\text{m} \pm 0.60 \mu\text{m}$, $34.1 \mu\text{m} \pm 0.67 \mu\text{m}$, and $39.9 \mu\text{m} \pm 0.58 \mu\text{m}$, respectively, where the uncertainties are standard uncertainties.

The measured values (but not the associated uncertainties) are depicted as red diamonds on the left panel of the figure below, which shows the cumulative distribution function of the particle size distribution listed in Table 1 of the certificate: the blue dots are measured values, and the blue, horizontal line segments issuing from the blue dots serve to indicate that this is a step function: that is, a function that is constant over (in this case fairly narrow) intervals, and then suddenly jumps to a higher value as d increases. The dots indicate that the function takes the higher value of $F(d)$ at each value of d where it jumps: in other words, the function is

right-continuous.



The corresponding certified values are $18.9 \mu\text{m} \pm 0.59 \mu\text{m}$, $26.7 \mu\text{m} \pm 0.51 \mu\text{m}$, $32.1 \mu\text{m} \pm 0.50 \mu\text{m}$, $37.4 \mu\text{m} \pm 0.63 \mu\text{m}$, and $43.3 \mu\text{m} \pm 0.48 \mu\text{m}$.

The goal of the calibration is to determine a function that, once applied to the diameter values measured by the user, calibrates them using the certified values. We will assume that a linear calibration function will suffice, which transforms a user-measured value d into $\alpha + \beta d$, where the “shift” α and the “gain” β have to be estimated using the calibration data. A.9 lists R code implementing this calibration, as described next.

The right panel of the figure above shows corresponding percentiles of the particle size distribution, as measured by the user along the horizontal axis, D_{SAMPLE} , and the corresponding certified value, D_{SRM} , along the vertical axis. Each point has an uncertainty in the horizontal direction, and another in the vertical direction. The short, blue line segments forming crosses over the red diamonds represent 1-sigma intervals (that is, coverage intervals with coverage probability approximately 68 %).

The green line fitted to these points on the right panel above, takes into account the uncertainties represented by the crosses. It was fitted using errors-in-variables regression [Fuller, 1987], by the method of maximum likelihood, assuming that measurement errors (in both directions) have Gaussian distributions with standard deviations equal to the corresponding standard uncertainties.

The yellow-green band around the fitted line is a 95 % coverage interval for the whole line. The band was computed by application of the parametric statistical bootstrap [Efron and Tibshirani, 1993], similarly to what is described by Guenther and Possolo [2011] for the uncertainty of analysis functions used to assign values to NIST gas mixture SRMs.

The fitted line has intercept $\hat{\alpha} = -1.30 \pm 1.41$ and slope $\hat{\beta} = 1.12 \pm 0.05$ (where the hats over the Greek letters indicate that they are estimates of the true intercept and slope), and the correlation between intercept and slope is -0.97 . These results suggest that α may very well actually be 0, and that the user’s measurement system underestimates the diameters by about 12 %.

The calibrated diameter corresponding to a user-measured diameter D is $D^* = \hat{\alpha} + \hat{\beta}D$. The standard uncertainty associated with the calibrated diameter is $u(D^*) = (u^2(\hat{\alpha}) + 2u(\hat{\alpha})u(\hat{\beta})rD + u^2(\hat{\beta})D^2)^{1/2}$, where $r = -0.97$ denotes the correlation coefficient between $\hat{\alpha}$ and $\hat{\beta}$.

For example, a user-measured diameter of $34.1 \mu\text{m}$ should become $-1.30 + (1.12 \times 34.1 \mu\text{m}) = 36.9 \mu\text{m}$, with standard uncertainty $(1.7^2 + 2 \times (1.7) \times (0.05) \times (-0.97) \times 34.1 + 0.05^2 \times 34.1^2)^{1/2} \mu\text{m} = 0.4 \mu\text{m}$.

This calculation does not include the uncertainty that may be associated with 34.1 μm . Taking this uncertainty into account, together with the uncertainties above, may most easily be done using a Monte Carlo method.

The small red dots shown on the left panel (each very close to, or on top of the corresponding large blue dot) are the calibrated versions of the red diamonds. And the width of each red rectangle represents the calibrated value plus or minus twice its standard uncertainty (the height of the rectangles has no meaning; it is there simply to facilitate the visualization of the uncertainties).

8.6 Some measurement results for scalar measurands include uncertainties expressed asymmetrically relative to the measured value. This may involve reporting the measured value x and qualifying it with “left” and “right” standard uncertainties, $u^-(x)$ and $u^+(x)$, or simply giving a 95 % coverage interval $(x - U_{95\%}^-(x), x + U_{95\%}^+(x))$ for the true value of the measurand such that $U_{95\%}^-(x)$ is different from $U_{95\%}^+(x)$.

8.7 Although uncommon among reference materials produced by NIST, such asymmetric expressions of uncertainty are very common in other fields, in particular in particle physics [Gavrilyuk et al., 2013] and in astronomy [Crowther et al., 2016; The Event Horizon Telescope Collaboration, 2019]. Asymmetrical uncertainties have also been reported in measurement science, for example by Yoshino et al. [1988], Nelson et al. [2015, Page 8657], and Hodges et al. [2019]. Accordingly, the topic of how to interpret and use such asymmetrical expressions of uncertainty has received appreciable attention, for example from Barlow [2003, 2004], Audi et al. [2017, Appendix A], and Possolo et al. [2019].

8.8 An approach commonly employed to propagate uncertainties expressed asymmetrically involves “symmetrizing” them and using the result as if the uncertainties in question had been originally expressed in the usual, symmetric form. The techniques that Possolo et al. [2019] describe, some of which are implemented in the *NIST Uncertainty Machine*, do not involve such simplification. Instead, they model the asymmetric uncertainties explicitly, using asymmetric probability distributions, and propagate these using a Monte Carlo method.

EXAMPLE 8.8.A NIST SRM 606 Trace Elements in Basalt Glass is one of the few reference materials produced by NIST where uncertainties are expressed asymmetrically around the certified values. The material is a glass obtained by melting the basalt of NIST SRM 688 Basalt Rock in the laboratory. The certified values are part of the following measurement results for the mass fractions of barium and strontium:

ELEMENT		w	$u(w)$	$w - U_{95\%}^-(w)$	$w + U_{95\%}^+(w)$	/ mg/kg
Barium	Ba	174	5	166	186	
Strontium	Sr	169	5	160	178	

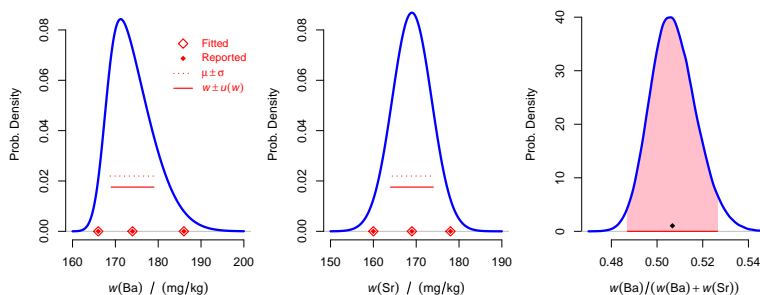
Now, consider the problem of estimating $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$, and evaluating the associated uncertainty. We will describe two approaches that lead to the same

solution: the first is developed in explicit detail and is implemented in R code listed in A.10; the second uses the *NIST Uncertainty Machine*, yet without describing the portions of its internal machinery that are involved in this case. Both approaches involve probabilistic models for the mass fractions of barium and of strontium that reproduce, with satisfactory accuracy, the information the certificate provides.

The first approach illustrates the use of the so-called *skew-normal distribution* [Azzalini and Capitanio, 2014] for this purpose. It should be noted that this model is not the best for all cases where asymmetric uncertainties must be modeled probabilistically, and it may even be inappropriate in some cases. However, both the manner of fitting whichever model is satisfactory for this purpose, and of propagating the uncertainty that it captures, are applicable generally.

The skew-normal distribution has three parameters: location ξ , scale $\omega > 0$, and shape α , real numbers all. The mean and standard deviation of the distribution can be written explicitly in terms of these parameters. However, finding the 2.5th and 97.5th percentiles that correspond to $w - U_{95\%}^-(w)$, and $w + U_{95\%}^+(w)$, involves numerical methods. All necessary computational facilities are available in R package `sn` [Azzalini, 2020].

For this illustration we chose values of ξ , ω , and α that minimize a weighted sum of squared differences between w , $\log u(w)$, $w - U_{95\%}^-(w)$, and $w + U_{95\%}^+(w)$, and their counterparts for the model.



In this case, the fitted distributions reproduce the measured values and the elements of the reported uncertainties very accurately. When they do not, the fit should aim to reproduce the endpoints of coverage intervals as closely as possible, possibly at the price of incurring a more substantial error at the measured value. This can be accomplished by giving greater weight to the errors in reproducing the endpoints of the coverage interval, $w - U_{95\%}^-(w)$ and $w + U_{95\%}^+(w)$, because the tails of the distribution influence the uncertainty propagation the most.

The next step is to draw large samples of the same size K from the fitted distributions, forming the corresponding ratios $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$, and then deriving from this sample of values of the ratio, an estimate for the ratio, and an expression of measurement uncertainty.

The rightmost panel of the figure above indicates the median of $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$ with the (blue) diamond near the horizontal axis, and the shaded (pink) region under the curve has area 0.95, chosen so that its footprint on the horizontal axis is centered at the estimate of the ratio. In conclusion, the estimate of $w(\text{Ba})/(w(\text{Ba}) + w(\text{Sr}))$ is 0.507 with expanded uncertainty for 95 % coverage of 0.020.

The second approach uses the *NIST Uncertainty Machine*, with the inputs shown below. Note that, in the line of boxes pertaining to Ba, first appears the measured value, $w(\text{Ba}) = 174 \text{ mg/kg}$, second $w - U_{95\%}^-(w) = 166 \text{ mg/kg}$, third $w + U_{95\%}^+(w) = 186 \text{ mg/kg}$, and fourth the coverage probability, 0.95, corresponding to the interval whose endpoints are 166 mg/kg and 186 mg/kg. Similarly for the line of boxes pertaining to Sr.

The screenshot shows the '1. Select Inputs & Choose Distributions' section of the NIST Uncertainty Machine. It includes a dropdown for 'Number of input quantities' set to 2, and input fields for 'Names of input quantities' with 'Ba' and 'Sr'. Below, there are rows for 'Ba' and 'Sr' with dropdown menus for distribution type (set to 'Asymmetric (Median, Left uncertainty, Right uncertainty, Co)') and input fields for median, lower uncertainty, upper uncertainty, and coverage probability (0.95). A 'Correlations' section is also visible. The '2. Choose Options' section shows 'Number of realizations of the output quantity' set to 1000000, 'Random number generator seed' set to 71, and 'Symmetrical coverage intervals' checked. The '3. Write the Definition of Output Quantity' section shows the R expression 'Ba/(Ba+Sr)'.

8.9 It is possible, and indeed quite easy to propagate the uncertainty associated with a qualitative property when the uncertainty is expressed quantitatively in terms of a probability distribution over the set of possible values of the property. The following example is a continuation of Example 6.7.c, which uses FASTQ-style quality scores extracted from a Solexa prb file distributed as part of the Bioconductor package ShortRead [Morgan et al., 2009], to illustrate how the corresponding uncertainties, expressed using locus-specific, discrete probability distributions, can be propagated to a derivative quantity. In due course it will be replaced by a similar example involving a NIST reference material.

EXAMPLE 8.9.A Consider the following result of sequencing a DNA fragment, where A denotes adenine, C cytosine, G guanine, and T thymine,

$$\theta = (\text{TTTTTATAATTGGTTAATCATTFTTTTTTAATTTTT}),$$

and suppose that we wish to compare it with a reference sequence

$$\tau = (\text{GGATTTTATTATAAATGGGTATACAATTTTAAAATTTTT}),$$

assumed known without uncertainty. These two sequences can be compared in terms of the Damerau-Levenshtein distance $D(\theta, \tau)$ between them [Damerau, 1964; Levenshtein, 1966].

This distance is the minimum number of insertions, deletions, or substitutions of a single character, or transpositions of two adjacent characters, that are needed to transform one string into the other. In this case, $D(\theta, \tau) = 13$ was computed using R function `stringdist` defined in the package of the same name [van der Loo, 2014].

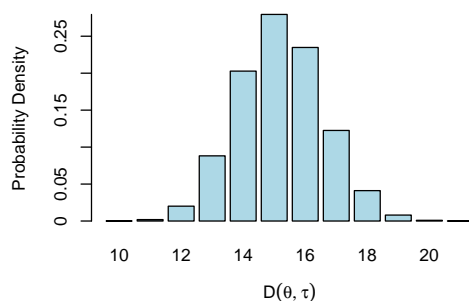
To characterize the uncertainty associated with $D(\theta, \tau)$, we need to propagate the uncertainty evaluations for the nucleobase identifications in θ , which are expressed in the form of discrete probability distributions, as already reviewed in Example 6.7.c, one for each of the 36 loci in θ , listed in the following table:

LOC	BASE	Pr(A)	Pr(C)	Pr(G)	Pr(T)	LOC	BASE	Pr(A)	Pr(C)	Pr(G)	Pr(T)
1	T	0.00	0.00	0.00	1.00	19	C	0.03	0.94	0.00	0.03
2	T	0.00	0.00	0.00	1.00	20	A	0.94	0.02	0.00	0.04
3	T	0.00	0.00	0.39	0.61	21	T	0.00	0.01	0.00	0.99
4	T	0.01	0.33	0.00	0.66	22	T	0.03	0.00	0.07	0.89
5	T	0.00	0.01	0.00	0.99	23	T	0.00	0.39	0.00	0.61
6	A	0.75	0.00	0.02	0.24	24	T	0.24	0.09	0.00	0.67
7	T	0.04	0.09	0.06	0.81	25	T	0.00	0.01	0.09	0.90
8	A	1.00	0.00	0.00	0.00	26	T	0.00	0.00	0.00	1.00
9	A	0.89	0.11	0.00	0.00	27	T	0.00	0.00	0.00	1.00
10	T	0.05	0.14	0.00	0.81	28	T	0.00	0.00	0.00	1.00
11	T	0.00	0.00	0.00	1.00	29	T	0.00	0.00	0.00	1.00
12	G	0.44	0.00	0.56	0.00	30	A	0.56	0.20	0.04	0.20
13	G	0.09	0.00	0.86	0.05	31	A	0.63	0.08	0.00	0.29
14	T	0.09	0.20	0.28	0.43	32	T	0.14	0.00	0.00	0.86
15	T	0.00	0.00	0.00	1.00	33	T	0.46	0.01	0.08	0.46
16	A	1.00	0.00	0.00	0.00	34	T	0.00	0.00	0.00	1.00
17	A	0.95	0.04	0.02	0.00	35	T	0.00	0.00	0.00	1.00
18	T	0.00	0.00	0.06	0.94	36	T	0.00	0.00	0.00	1.00

To characterize the associated uncertainty, employ a Monte Carlo method:

1. Select a suitably large sample size m , and then repeat the next step m times, for $i = 1, \dots, m$;
2. For each locus table, draw a letter from $\{A, C, G, T\}$ using the probabilities listed for that locus, to obtain a string θ_i^* whose characters are the letters drawn for each of the thirty-six loci;
3. The distances $D(\theta_1^*, \tau), \dots, D(\theta_m^*, \tau)$, between these sampled sequences and the target sequence, are a sample from the distribution of $D(\theta, \tau)$.

The following histogram summarizes a sample of $m = 10^5$ values of this distance, obtained as just described. It shows that $D(\theta, \tau) = 15$ is the value with highest probability, not the value, 13, that was actually observed. The sample has average 15.2 and standard deviation $u(D(\theta, \tau)) = 1.4$, and a 95 % coverage interval for $D(\theta, \tau)$ ranges from 13 to 18.



8.10 For some reference materials, not only does the certificate list certified values and uncertainties, but it also provides the means to characterize the underlying probability distribution. For example, the uncertainty associated with the certified amount concentration of 25-hydroxyvitamin D₃ can be described by a Gaussian distribution with mean 71.8 nmol/L and standard deviation 1.35 nmol/L. The probability distribution may also be characterized by providing a large sample drawn from the relevant distribution, which may be used in Monte Carlo procedures for uncertainty propagation.

EXAMPLE 8.10.A The certificate of NIST SRM 2780a Hard Rock Mine Waste offers certified values for the mass fractions of many chemical elements, and the associated uncertainties are expressed as asymmetric uncertainty intervals.

The certificate also includes a pointer to a NIST web site with supplementary information in the form of files containing samples drawn from the probability distributions that describe the uncertainty surrounding these mass fractions. These riches afford the user of this material two alternative approaches to propagate the reported uncertainties.

One approach uses only the information in the certificate: the certified value, which is the median of the probability distribution that characterizes the reported uncertainty, and the endpoints of the uncertainty interval, which are the 2.5th and 97.5th percentiles of the same distribution.

One would then do as described in Example 8.8.A, or, more simply still, use the modeling option Asymmetric (Median, Left uncertainty, Right uncertainty, Coverage probability, under 1. *Select Inputs & Choose Distributions*, in the *NIST Uncertainty Machine*.

The other approach involves no modeling, relying entirely on the supplementary information, and employs a Monte Carlo method for uncertainty evaluation. This can be done using the *NIST Uncertainty Machine*, as the certificate also explains, because this application allows the user to assign a distribution to an input quantity in the form of a sample.

These two ratios of amounts of substance are used in Pearce Element Ratio (PER) analysis [Pearce, 1968] in geochemistry: $(n(\text{Na}) + n(\text{K}))/n(\text{Ti})$ and $n(\text{Al})/n(\text{Ti})$, where $n(\text{Na})$ denotes the amount of sodium (expressed in moles), and similarly for the other elements. Stanley [2020] shows how a plot of pairs of values of these ratios can be used to characterize the hydrothermal alteration of alkali feldspar-bearing rocks.

Let $A_r(\text{Al}) = 26.981\,538\,4(3)$ and $A_r(\text{Ti}) = 47.867(1)$ denote the standard atomic weights of aluminum and titanium (from CIAAW, *Atomic weights of the elements 2019*, www.ciaaw.org). The associated uncertainties may be modeled using rectangular (or uniform) distributions [Possolo et al., 2018b]: for example, the rectangular distribution concentrated on $[47.866, 47.868]$ for titanium.

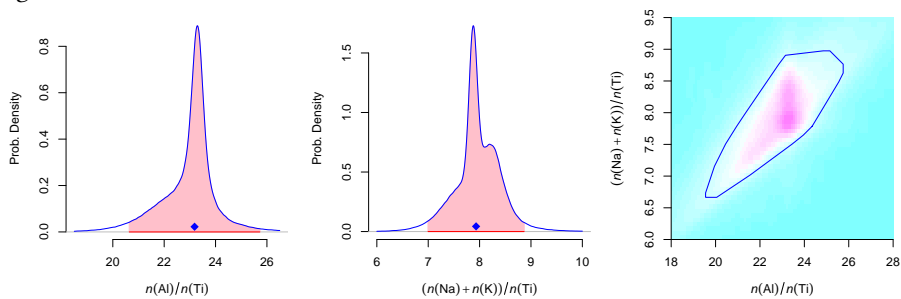
These are the steps needed to use the supplementary information available for this material to evaluate the uncertainty of $n(\text{Al})/n(\text{Ti})$, based on a sample of large size M :

- (1) Draw M values from the rectangular distributions used to model the uncertainties of the atomic weights of aluminum and titanium;
- (2) Draw M values, uniformly at random, with replacement, from the supplementary information file for aluminum in SRM 2780a (which has 50 000 values), and do likewise for titanium;
- (3) Convert each value of the mass fraction of aluminum, and the corresponding value of the mass fraction of titanium, into amounts per unit of mass of the material, so that the ratio of these quantities is the ratio of amount fractions.

The conversion for titanium is done as follows, supposing that the sampled value of the mass fraction is 0.643 cg/g, and the corresponding sampled value of the atomic weight is 47.867: since

this says that 1 g of material has 0.643 cg of titanium, and that the mass of 1 mol of titanium is 47.867 g, we conclude that there are $0.643/47.867$ cmol of titanium per gram of material. A similar calculation for aluminum will give us the amount of aluminum per gram of material. The ratio between these two amounts per gram is the amount fraction we seek.

These are the probability densities of $n(\text{Al})/n(\text{Ti})$ and $(n(\text{Na}) + n(\text{K}))/n(\text{Ti})$, with 95 % uncertainty intervals centered at their medians, and also the density of their bivariate, joint distribution, where the blue polygon outlines a 95 % confidence region.



References

- A. Agresti. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Hoboken, NJ, 3rd edition, 2019. [21]
- G. Audi, F.G. Kondev, M. Wang, W. J. Huang, and S. Naimi. The NUBASE2016 evaluation of nuclear properties. *Chinese Physics C*, 41(3):030001–1–138, March 2017. doi: 10.1088/1674-1137/41/3/030001. [39]
- A. Azzalini. *The R package sn: The Skew-Normal and Related Distributions such as the Skew-t (version 1.6-1)*. Università di Padova, Italia, 2020. URL <http://azzalini.stat.unipd.it/SN>. [40]
- A. Azzalini and A. Capitanio. *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge, UK, 2014. [40]
- R. Barlow. Asymmetric errors. In *PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pages 250–255, Menlo Park, CA, September 8th–11th 2003. SLAC National Accelerator Laboratory. URL www.slac.stanford.edu/econf/C030908/proceedings.html. [39]
- R. Barlow. Asymmetric Statistical Errors. *ArXiv Physics e-prints*, June 2004. [39]
- C. R. Beauchamp, J. E. Camara, J. Carney, S. J. Choquette, K. D. Cole, P. C. DeRose, D. L. Duewer, M. S. Epstein, M. C. Kline, K. A. Lippa, E. Lucon, K. W. Phinney, A. Possolo, K. E. Sharpless, J. R. Sieber, B. Toman, M. R. Winchester, and D. Windover. *Metrological Tools for the Reference Materials and Reference Instruments of the NIST Materials Measurement Laboratory*. NIST Special Publication 260-136 (2020 Edition). National Institute of Standards and Technology, Gaithersburg, MD, 2020. doi: 10.6028/NIST.SP.260-136-2020. [6, 8, 11, 34]
- T. A. Butler and J. L. Molloy. Preparation of candidate SRM1641e mercury in water. Report of Analysis 646.01-14-025, National Institute of Standards and Technology, Material Measurement Laboratory, Chemical Sciences Division, Gaithersburg, MD, June 2014. [20]
- N. R. Campbell. *Physics: The Elements*. Cambridge University Press, London, UK, 1920. [13]
- R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, Illinois, 2nd edition, 1962. [14]
- P. J. A. Cock, C. J. Fields, N. Goto, M. L. Heuer, and P. M. Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38:1767–1771, 2009. doi: 10.1093/nar/gkp1137. [22]
- W. J. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, NY, 3rd edition, 1999. [31]

- P. A. Crowther, S. M. Caballero-Nieves, K. A. Bostroem, J. Maíz Apellániz, F. R. N. Schneider, N. R. Walborn, C. R. Angus, I. Brott, A. Bonanos, A. de Koter, S. E. de Mink, C. J. Evans, G. Gräfener, A. Herrero, I. D. Howarth, N. Langer, D. J. Lennon, J. Puls, H. Sana, and J. S. Vink. The R136 star cluster dissected with Hubble Space Telescope/STIS. I. Far-ultraviolet spectroscopic census and the origin of He ii $\lambda 1640$ in young star clusters. *Monthly Notices of the Royal Astronomical Society*, 458(1):624–659, 2016. doi: 10.1093/mnras/stw273. [39]
- F. J. Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176, March 1964. doi: 10.1145/363958.363994. [41]
- M. H. DeGroot. *Probability and Statistics*. Addison-Wesley, Reading, MA, 2nd edition, 1986. ISBN 978-0-201-11366-2. [13]
- M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Addison-Wesley, Boston, MA, 4th edition, 2012. [15, 37]
- D. Diez, M. Çetinkaya-Rundel, and C. D. Barr. *OpenIntro Statistics*. Center for Open Education, University of Minnesota, Minneapolis, MN, Fourth edition, 2019. ISBN 978-1943450077. URL openintro.org/os. [13]
- B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, London, UK, 1993. [38]
- H. Emons, A. Fajgelj, A. M. H. van der Veen, and R. Watters. New definitions on reference materials. *Accreditation and Quality Assurance*, 10:576–578, 2006. doi: 10.1007/s00769-006-0089-9. [11]
- B. Ewing and P. Green. Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. *Genome Research*, 8:186–194, 1998. [23]
- B. Ewing, L. Hillier, M. C. Wendl, and P. Green. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. *Genome Research*, 8:175–185, 1998. [23]
- D. Freedman, R. Pisani, and R. Purves. *Statistics*. W. W. Norton & Company, New York, NY, 4th edition, 2007. ISBN 978-0-393-92972-0. [13, 16, 35]
- W. A. Fuller. *Measurement Error Models*. John Wiley & Sons, New York, NY, 1987. [38]
- C. F. Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Frid. Perthes and I. H. Besser, Hamburg, Germany, 1809. [16]
- Y. M. Gavriluk, A. M. Gangapshev, V. V. Kazalov, V. V. Kuzminov, S. I. Panasenkov, and S. S. Ratkevich. Indications of $2\nu 2K$ capture in ^{78}Kr . *Physical Review C*, 87:035501, 2013. doi: 10.1103/physrevc.87.035501. [39]
- F. R. Guenther and A. Possolo. Calibration and uncertainty assessment for certified reference gas mixtures. *Analytical and Bioanalytical Chemistry*, 399:489–500, 2011. doi: 10.1007/s00216-010-4379-z. [31, 38]

- A. Hájek. “Mises redux” — redux: Fifteen arguments against finite frequentism. *Erkenntnis*, 45(2–3):209–227, November 1996. [14]
- A. Hájek. Interpretations of probability. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, California, 2007. URL plato.stanford.edu/archives/win2007/entries/probability-interpret/. [13]
- A. Hájek. Fifteen arguments against hypothetical frequentism. *Erkenntnis*, 70:211–235, 2009. [14]
- J. Hodges, J. Viallon, P. J. Brewer, B. J. Drouin, V. Gorshelev, C. Janssen, S. Lee, A. Possolo, M.-A. H. Smith, J. Walden, and R. Wielgosz. Recommendation of a consensus value of the ozone absorption cross-section at 253.65 nm based on literature review. *Metrologia*, 53(3):034001, 2019. doi: 10.1088/1681-7575/ab0bdd. [39]
- J. L. Hodges and E. L. Lehmann. *Basic Concepts of Probability and Statistics*, volume 48 of *Classics in Applied Mathematics*. Society for Industrial Mathematics, Philadelphia, PA, Second edition, 2005. ISBN 978-0816240043. [13]
- ISO. *Gas analysis — Comparison methods for determining and checking the composition of calibration gas mixtures*. International Organization for Standardization (ISO), Geneva, Switzerland, 2001. International Standard ISO 6143:2001(E). [31]
- ISO. *Reference materials — Selected terms and definitions*. International Organization for Standardization (ISO), Geneva, Switzerland, Third edition, February 2015. ISO Guide 30:2015(E). [11]
- Joint Committee for Guides in Metrology. *Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method*. International Bureau of Weights and Measures (BIPM), Sèvres, France, 2008. URL www.bipm.org/en/publications/guides/gum.html. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, JCGM 101:2008. [17, 32]
- Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. International Bureau of Weights and Measures (BIPM), Sèvres, France, 2008. URL www.bipm.org/en/publications/guides/gum.html. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, JCGM 100:2008, GUM 1995 with minor corrections. [10, 11]
- A. La Caze. Frequentism. In A. Hájek and C. Hitchcock, editors, *The Oxford Handbook of Probability and Philosophy*, chapter 25. Oxford University Press, Oxford, UK, 2016. [14]
- T. Lafarge and A. Possolo. The NIST Uncertainty Machine. *NCSLI Measure Journal of Measurement Science*, 10(3):20–27, September 2015. [20]
- V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics — Doklady*, 10(8):707–710, February 1966. [41]

- M. Li, J. Tan, M. J. Tarlov, and M. R. Zachariah. Absolute quantification method for protein concentration. *Analytical Chemistry*, 86:12130–12137, November 2014. doi: 10.1021/ac5030123. [26, 27]
- D. J. Lipman and W. R. Pearson. Rapid and sensitive protein similarity searches. *Science*, 227:1435–1441, March 1985. doi: 10.1126/science.2983426. [22]
- W. May, R. Parris, C. Beck, J. Fassett, R. Greenberg, F. Guenther, G. Kramer, S. Wise, T. Gills, J. Colbert, R. Gettings, and B. MacDonald. *Definitions of Terms and Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements*. National Institute of Standards and Technology, Gaithersburg, Maryland, January 2000. NIST Special Publication 260-136. [34]
- M. Morgan, S. Anders, M. Lawrence, P. Aboyou, H. Pagès, and R. Gentleman. ShortRead: a Bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25:2607–2608, 2009. doi: 10.1093/bioinformatics/btp450. [22, 23, 41]
- F. Mosteller and J. W. Tukey. *Data Analysis and Regression*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1977. [15]
- M. A. Nelson, M. Bedner, B. E. Lang, B. Toman, and K. A. Lippa. Metrological approaches to organic chemical purity: primary reference materials for vitamin D metabolites. *Analytical and Bioanalytical Chemistry*, 407:8557–8569, November 2015. doi: 10.1007/s00216-015-9013-7. [39]
- M. A. Nelson, J. F. Waters, B. Toman, B. E. Lang, A. Rück, K. Breitruick, M. Obkircher, A. Windust, and K. A. Lippa. A new realization of SI for organic chemical measurement: NIST PS1 primary standard for quantitative NMR (Benzoic Acid). *Analytical Chemistry*, 90:10510–10517, September 2018. doi: 10.1021/acs.analchem.8b02575. [29]
- NIST. *NIST Quality Manual for Measurement Services*. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, Maryland, December 2019. URL <https://www.nist.gov/nist-quality-system/>. NIST QM-I Version 11. [11, 12]
- A. O’Hagan. Eliciting and using expert knowledge in metrology. *Metrologia*, 51(4): S237–S244, 2014. doi: 10.1088/0026-1394/51/4/S237. [13]
- R. C. Paule and J. Mandel. *Analysis of Interlaboratory Measurements on the Vapor Pressure of Gold (Certification of Standard Reference Material 745)*. National Bureau of Standards, Washington, DC, January 1970. Special Publication 260-19. [11]
- T. H. Pearce. A contribution to the theory of variation diagrams. *Contributions to Mineralogy and Petrology*, 19:142–157, 1968. doi: 10.1007/BF00635485. [43]
- K. R. Popper. *A World of Propensities — Two New Views on Causality*. Thoemmes Press, Bristol, UK, 1990. [14]

- A. Possolo. *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. National Institute of Standards and Technology, Gaithersburg, MD, 2015. URL <http://dx.doi.org/10.6028/NIST.TN.1900>. NIST Technical Note 1900. [11]
- A. Possolo and B. Toman. *Tutorial for metrologists on the probabilistic and statistical apparatus underlying the GUM and related documents*. National Institute of Standards and Technology, Gaithersburg, MD, November 2011. doi: 10.13140/RG.2.1.2256.8482. URL www.itl.nist.gov/div898/possolo/TutorialWEBServer/TutorialMetrologists2011Nov09.xht. [13]
- A. Possolo, O. Bodnar, T. A. Butler, J. L. Molloy, and M. R. Winchester. Value assignment and uncertainty evaluation in single-element reference solutions. *Metrologia*, 55(3): 404–413, 2018a. doi: 10.1088/1681-7575/aabd57. [25]
- A. Possolo, A. M. H. van der Veen, J. Meija, and D. B. Hibbert. Interpreting and propagating the uncertainty of the standard atomic weights. *Pure and Applied Chemistry*, 90(2):395–424, 2018b. doi: 10.1515/pac-2016-0402. IUPAC Technical Report. [17, 35, 43]
- A. Possolo, C. Merkatas, and O. Bodnar. Asymmetrical uncertainties. *Metrologia*, 56(4): 045009, 2019. doi: 10.1088/1681-7575/ab2a8d. [17, 39]
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>. [6, 16]
- G. C. Rhoderick, D. R. Kitzis, M. E. Kelley, W. R. Miller, B. D. Hall, E. J. Dlugokencky, P. P. Tans, A. Possolo, and J. Carney. Development of a northern continental air standard reference material. *Analytical Chemistry*, 88(6):3376–3385, 2016. doi: 10.1021/acs.analchem.6b00123. [20]
- F. Sanger and A. R. Coulson. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3):441–448, May 1975. doi: 10.1016/0022-2836(75)90213-2. [23]
- F. Sanger, S. Nicklen, and A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, December 1977. doi: 10.1073/pnas.74.12.5463. [23]
- F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, December 1946. [18]
- K. E. Sharpless, K. A. Lippa, D. L. Duewer, and A. L. Rukhin. *The ABCs of Using Standard Reference Materials in the Analysis of Foods and Dietary Supplements: A Practical Guide*. NIST Special Publication 260-181r1. National Institute of Standards and Technology, Gaithersburg, MD, 2015. doi: 10.6028/NIST.SP.260-181r1. [11]

- D. A. Skoog, D. M. West, F. J. Holler, and S. R. Crouch. *Fundamentals of Analytical Chemistry*. Brooks/Cole, Belmont, CA, 9th edition, 2014. ISBN 978-0-495-55828-6. [18]
- L. T. Sniegowski and J. R. Moody. Determination of serum and blood densities. *Analytical Chemistry*, 51(9):1577–1578, August 1979. doi: 10.1021/ac50045a052. [35]
- J. D. Splett, C. N. McCowan, H. K. Iyer, and C.-M. Wang. *NIST Recommended Practice Guide: Computing Uncertainty for Charpy Impact Machine Test Results*. NIST Special Publication 960-18. National Institute of Standards and Technology, Gaithersburg, MD, 2008a. [5]
- J. D. Splett, C. N. McCowan, and C.-M. Wang. *Charpy Machine Verification: Limits and Uncertainty*. NIST Special Publication 260-171. National Institute of Standards and Technology, Gaithersburg, MD, 2008b. URL nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication260-171.pdf. [5]
- C. R. Stanley. Molar element ratio analysis of lithochemical data: a toolbox for use in mineral exploration and mining. *Geochemistry: Exploration, Environment, Analysis*, 20: 233, 2020. doi: 10.1144/geochem2019-033. [43]
- B. N. Taylor and C. E. Kuyatt. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST Technical Note 1297. National Institute of Standards and Technology, Gaithersburg, MD, 1994. URL <https://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf>. [11, 32, 35]
- The Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope results. VI. The shadow and mass of the central black hole. *The Astrophysical Journal Letters*, 875(1):L6, April 2019. doi: 10.3847/2041-8213/ab1141. [39]
- M. Thompson and S. L. R. Ellison. Dark uncertainty. *Accreditation and Quality Assurance*, 16:483–487, October 2011. doi: 10.1007/s00769-011-0803-0. [25]
- J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, 1977. [16]
- M.P.J. van der Loo. The stringdist package for approximate string matching. *The R Journal*, 6, 2014. URL CRAN.R-project.org/package=stringdist. Accepted for publication. [41]
- K. L. Vink, W. Schuurman, and R. van Gansewinkel. Use of the caffeine reagent in direct spectrophotometry of bilirubin. *Clinical Chemistry*, 32(7):1389–1393, July 1986. doi: 10.1093/clinchem/32.7.1389. [20]
- K. L. Vink, W. Schuurman, and R. van Gansewinkel. Direct spectrophotometry of bilirubin in serum of the newborn, with use of caffeine reagent. *Clinical Chemistry*, 34(1):67–70, January 1988. doi: 10.1093/clinchem/34.1.67. [20]
- B. L. Welch. The generalization of ‘Student’s’ problem when several different population variances are involved. *Biometrika*, 34:28–35, January 1947. [18]

K. Yoshino, D.E. Freeman, J.R. Esmond, and W.H. Parkinson. Absolute absorption cross-section measurements of ozone in the wavelength region 238-335 nm and the temperature dependence. *Planetary and Space Science*, 36(4):395–398, 1988. doi: 10.1016/0032-0633(88)90127-4. [\[39\]](#)

Appendix: R Code

A.1 R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, including the Linux variants, and Microsoft Windows and Apple's MacOS.

Visit <https://www.r-project.org/>, click on CRAN (top left, under Download), and then select a mirror near you. Next, select the appropriate download option under Download and Install R, and install R.

Visit <https://www.rstudio.com/products/rstudio/download/> and download and install the version of RStudio suitable for your operating system. Snippets of R code scattered throughout this document, including in this Appendix, may be copied and pasted into RStudio, and executed there.

Go to <https://cran.r-project.org/manuals.html> and read *An Introduction to R*: HTML, PDF, and EPUB versions are available.

Required R Packages:

```
if (!require(metafor)) {install.packages("metafor")}
if (!require(metRology)) {install.packages("metRology")}
if (!require(nloptr)) {install.packages("nloptr")}
if (!require(sn)) {install.packages("sn")}
if (!require(robustbase)) {install.packages("robustbase")}
```

A.2 Example 7.2.A

NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)

```
c.NIST = 39.4 ## NIST certified value
uc.NIST = 0.8/2 ## NIST standard uncertainty

c.1 = 34.85 ## First user's measured value (ng/mL)
uc.1 = 2.45/2 ## First user's standard uncertainty
z.1 = (c.1 - c.NIST) / sqrt(uc.1^2 + uc.NIST^2) ## z-test criterion
2*pnorm(-abs(z.1)) ## p-value

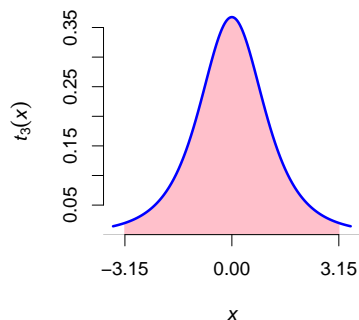
c.2 = 37.45 ## Second user's measured malue
uc.2 = 1.75/2 ## Second user's standard uncertainty
z.2 = (c.2 - c.NIST) / sqrt(uc.2^2 + uc.NIST^2) ## z-test criterion
2*pnorm(-abs(z.2)) ## p-value
```

A.3 Example 7.2.B

NIST SRM 3280 Multivitamin/Multielement Tablets

Given a coverage factor k and a coverage probability, the R function `kInverter` computes the number of degrees of freedom of the corresponding Student's t distribution.

In the example below, following the definition of the function, both the coverage factor (3.15) and the coverage probability (95 %) are specified as inputs. The function then outputs $\nu = 3$. This means that the area under the curve representing the probability density of Student's t with 3 degrees of freedom, and between -3.15 and $+3.15$ on the horizontal axis, amounts to 95 % of the total area under the curve.



```
kInverter = function (k, coverage=0.95)
{
  if (k < qnorm((1+coverage)/2)) {
    cat(paste0("WARNING: Gaussian measurement error model",
              " may not be appropriate,\n",
              "          and number of degrees of freedom",
              " may not be meaningful\n"))
    return(Inf)
  } else {
    kEqn = function (nu, kFactor, gamma) { qt(gamma, df=nu) - kFactor }
    kEqn.root = try(uniroot(f=kEqn, lower=sqrt(.Machine$double.eps),
                          upper=.Machine$integer.max,
                          kFactor=k, gamma=(1+coverage)/2))
    if (class(kEqn.root) != "try-error") { return(kEqn.root$root) }
    else { cat(paste0("kInverter could not find number of ",
                     "degrees of freedom corresponding to ",
                     k, "\n")); return(NA) }
  }
}

kInverter(k=3.15, coverage=0.95)
```

A.4 Example 7.2.c

NIST SRM 3161a Tin (Sn) Standard Solution (Lot No. 140917)

```
w.GR = 10.00007 ## Gravimetric determination
uw.GR = 0.00031
w.ICP = 10.02173 ## ICP-OES determination
uw.ICP = 0.00264

## DerSimonian-Laird estimate of dark uncertainty (tau)
sqrt(rma(yi=c(w.GR, w.ICP), sei=c(uw.GR, uw.ICP), method="DL")$tau2)
```

A.5 Example 7.2.D

NIST SRM 927e Bovine Serum Albumin (7 % Solution)

The following code yields p -value of 0.12, the same as the calculations in Example 7.2.D. However, some of the intermediate quantities differ, for example ν , owing to not having performed any rounding here, until the last step.

```
## Certified value and associated standard uncertainty
c.cert = 67.38 ## g/L
uc.cert = 1.38/2
```

```

nu.cert = kInverter(k=2, coverage=0.95)

## User results
c.user = mean(c(66.3, 67.6, 63.6, 66.5, 64.9))
uc.user = sd(c(66.3, 67.6, 63.6, 66.5, 64.9))/sqrt(5)
nu.user = 5-1

## Test criterion
t = (c.user - c.cert) / sqrt(uc.cert^2 + uc.user^2)
## Effective number of degrees of freedom
nu = welch.satterthwaite(ui=c(uc.user, uc.cert), df=c(nu.user, nu.cert))

## p-value
round(2*pt(-abs(t), df=nu), 2)

```

A.6 Example 7.3.A NIST SRM 606 Trace Elements in Basalt Glass

```

w.606 = 315 ## Mass fraction of Cr in SRM 606
uw.606 = 9 ## Std. unc.
Lwr95.606 = 302 ## Lower endpoint of 95 % CI
Upr95.606 = 336 ## Upper endpoint of 95 % CI

w.688 = 332 ## Mass fraction of Cr in SRM 606
uw.688 = 9 ## Std. unc.

lognormalMatchmaker =
  function (theta, w, uw, Lwr, Upr, gamma=0.95, weights=rep(1,4),
           tol=sqrt(.Machine$double.eps))
  {
    if (any(theta < tol)) {return(Inf)}
    } else {
      mu = theta[1]; sigma = theta[2]
      E = exp(mu + sigma^2/2)
      S = sqrt((exp(sigma^2)-1)*exp(2*mu+sigma^2))
      QLwr = qlnorm((1-gamma)/2, meanlog=mu, sdlog=sigma)
      QUpr = qlnorm((1+gamma)/2, meanlog=mu, sdlog=sigma)
      return(weights[1]*(w-E)^2 + weights[2]*(uw/w - S/E)^2
             + weights[3]*(Lwr-QLwr)^2 + weights[4]*(Upr-QUpr)^2)
    }
  }

o.606 = optim(par=c(log(w.606), uw.606/w.606), fn=lognormalMatchmaker,
             w=w.606, uw=uw.606, Lwr=Lwr95.606, Upr=Upr95.606,
             weights=c(1,1,4,4))
muHAT.606 = o.606$par[1]
sigmaHAT.606 = o.606$par[2]
cbind(obs=c(w=w.606, uw=uw.606, Lwr95=Lwr95.606, Upr95=Upr95.606),
      model=round(c(exp(muHAT.606 + sigmaHAT.606^2/2),
                    sqrt((exp(sigmaHAT.606^2)-1)*exp(2*muHAT.606+sigmaHAT.606^2)),
                    qlnorm(0.025, meanlog=muHAT.606, sdlog=sigmaHAT.606),
                    qlnorm(0.975, meanlog=muHAT.606, sdlog=sigmaHAT.606)), 0))

muHAT.688 = log(w.688 / sqrt((uw.688/w.688)^2+1))
sigmaHAT.688 = sqrt(log((uw.688/w.688)^2 + 1))
cbind(obs=c(w=w.688, uw=uw.688),
      model=round(c(exp(muHAT.688 + sigmaHAT.688^2/2),
                    sqrt((exp(sigmaHAT.688^2)-1)*exp(2*muHAT.688+sigmaHAT.688^2))), 0))

## Comparing estimates of parameters of lognormal models

```

```

z = (muHAT.606 - muHAT.688) / sqrt(sigmaHAT.606^2 + sigmaHAT.688^2)
2*pnorm(-abs(z))

## Comparing measured values using certified values and associated
## standard uncertainties, considering that the standard uncertainty
## of the (base e) logarithm of a measured value is approximately
## equal to its relative standard uncertainty
z = (log(w.606)-log(w.688)) / sqrt((uw.606/w.606)^2 + (uw.688/w.688)^2)
2*pnorm(-abs(z))

```

A.7 Example 7.4.A

NIST SRM 1003c Glass Beads (Particle Size Distribution)

```

p = seq(from=0.05, to=0.95, by=0.05)
B1.LLS = c(18.8, 22.1, 24.2, 25.7, 27.1, 28.2, 29.2, 30.3, 31.2,
           32.1, 33.0, 34.0, 35.0, 35.9, 37.1, 38.2, 39.4, 41.2, 43.1)
B1.ESZ = c(19.3, 21.9, 23.8, 25.2, 26.6, 27.8, 28.9, 30.0, 31.1,
           32.2, 33.3, 34.4, 35.5, 36.6, 37.8, 39.0, 40.2, 41.7, 43.5)

x = rnorm(length(p))
e = ecdf(x)
assign(x="x", value=B1.LLS, envir=environment(e))
assign(x="y", value=p, envir=environment(e))
B1.LLS.ecdf = e

x = rnorm(length(p))
e = ecdf(x)
assign(x="x", value=B1.ESZ, envir=environment(e))
assign(x="y", value=p, envir=environment(e))
B1.ESZ.ecdf = e

## Kolmogorov-Smirnov Test
D = c(B1.LLS, B1.ESZ)
testStatistic = max(abs(B1.LLS.ecdf(D)-B1.ESZ.ecdf(D)))

pkstwo = function(x, tol = 1e-06)
{
  if (is.numeric(x)) {x = as.double(x)}
  else {stop("argument 'x' must be numeric")}
  p = rep(0, length(x))
  p[is.na(x)] = NA
  IND = which(!is.na(x) & (x > 0))
  if (length(IND)) {
    p[IND] = .Call(stats:::C_pKS2, p = x[IND], tol) }
  return(p)
}

n = 1000 ## No. of particles counted
1 - pkstwo(sqrt(n) * testStatistic)

```

A.8 Example 8.4.A

NIST SRM 2973 Vitamin D Metabolites in Frozen Human Serum (High Level)

```

## Density
D = c(1.02230, 1.02230, 1.02226) ## g/mL
D.x = mean(D)
D.u = sd(D)/sqrt(3)

```

```

## Molecular Weight -- C27H44O2
C.Lwr = 12.0096; C.Upr = 12.0116
H.Lwr = 1.00784; H.Upr = 1.00811
O.Lwr = 15.99903; O.Upr = 15.99977

C.x = mean(c(C.Upr, C.Lwr))
H.x = mean(c(H.Upr, H.Lwr))
O.x = mean(c(O.Upr, O.Lwr))
C.u = abs(diff(c(C.Upr, C.Lwr)))/sqrt(12)
H.u = abs(diff(c(H.Upr, H.Lwr)))/sqrt(12)
O.u = abs(diff(c(O.Upr, O.Lwr)))/sqrt(12)

M.x = 27*(C.Lwr+C.Upr)/2 + 44*(H.Lwr+H.Upr)/2 + 2*(O.Lwr+O.Upr)/2
M.u = sqrt((27^2)*(C.Upr-C.Lwr)^2/12 + (44^2)*(H.Upr-H.Lwr)^2/12 +
(2^2)*(O.Upr-O.Lwr)^2/12)

## Mass fraction
w.x = 38.6; w.u = 0.4

## Amount-of-substance concentration
c.x = 1000 * w.x * D.x / M.x
c.u = c.x * sqrt( (w.u/w.x)^2 + (D.u/D.x)^2 + (M.u/M.x)^2 )
c(c.x, c.u)

```

A.9 Example 8.5.A NIST SRM 1003c Glass Beads (Particle Size Distribution)

```

## Certified particle size distribution
D = c(18.9, 21.8, 23.7, 25.3, 26.7, 27.9, 29.0, 30.1, 31.1, 32.1,
33.1, 34.1, 35.2, 36.2, 37.4, 38.6, 39.8, 41.4, 43.3)
D.u = c(0.59, 0.50, 0.50, 0.51, 0.51, 0.50, 0.49, 0.49, 0.49, 0.50,
0.52, 0.55, 0.58, 0.62, 0.63, 0.62, 0.64, 0.54, 0.48)
D.p = seq(from=0.05, to=0.95, by=0.05)

e = ecdf(D)
assign(x="x", value=D, envir=environment(e))
assign(x="y", value=D.p, envir=environment(e))
D.ecdf = e

D.xy = sortedXyData(D, D.p)
DS.p = c(0.05, 0.25, 0.50, 0.75, 0.95)
D05 = D[which.min(abs(D.p-0.05))]; D05.u = D.u[which.min(abs(D.p-0.05))]
D25 = D[which.min(abs(D.p-0.25))]; D25.u = D.u[which.min(abs(D.p-0.25))]
D50 = D[which.min(abs(D.p-0.50))]; D50.u = D.u[which.min(abs(D.p-0.50))]
D75 = D[which.min(abs(D.p-0.75))]; D75.u = D.u[which.min(abs(D.p-0.75))]
D95 = D[which.min(abs(D.p-0.95))]; D95.u = D.u[which.min(abs(D.p-0.95))]
DD = c(D05, D25, D50, D75, D95)
DD.u = c(D05.u, D25.u, D50.u, D75.u, D95.u)

## Percentiles determined by user
DS = c(18.1, 25.0, 29.5, 34.1, 39.9)
DS.u = c(0.73, 0.65, 0.60, 0.67, 0.58)

## Calibration function
ss = function(theta, x, ux, y, uy)
{
  alpha = theta[1]
  beta = theta[2]
  xi = theta[-(1:2)]

```

```

eta = alpha + beta*xi
return(-1*sum(dnorm(x, mean=xi, sd=ux, log=TRUE) +
             dnorm(y, mean=eta, sd=uy, log=TRUE)))
}

## The slope and intercept obtained using "lm" next,
## serve only as initial values for the nonlinear optimization
## that will produce the estimates alphaHAT and betaHAT below
DD.lm = summary(lm(DD~DS))

ab = coef(DD.lm)[,"Estimate"]
uab = coef(DD.lm)[,"Std. Error"]
oNL = nloptr(x0=c(ab, DS), eval_f=ss,
            opts=list(algorithm="NLOPT_LN_SBPLX",
                    "xtol_rel"=1.0e-7, maxeval=1e6, maxtime=-1),
            x=DS, ux=DS.u, y=DD, uy=DD.u)

alphaHAT = oNL$solution[1]
betaHAT = oNL$solution[2]
c(alphaHAT, betaHAT)
xiHAT = oNL$solution[-(1:2)]
etaHAT = alphaHAT + betaHAT*xiHAT

## Monte Carlo evaluation of uncertainty of slope and intercept of
## calibration function, and of correlation between estimates of
## slope and intercept. WARNING: The execution of the following code
## will take considerable time
K = 50000
abB = array(dim=c(K,2))
nD = length(DD)
for (k in 1:K)
{
  if (k %% 100 == 0) {cat(k, "of", K, "\n")}
  DSB = rnorm(nD, mean=xiHAT, sd=DS.u)
  DDB = rnorm(nD, mean=etaHAT, sd=DD.u)
  oNLB = nloptr(x0=c(alphaHAT, betaHAT, xiHAT),
              eval_f=ss,
              opts=list(algorithm="NLOPT_LN_SBPLX",
                      "xtol_rel"=1.0e-7, maxeval=1e6, maxtime=-1),
              x=DSB, ux=DS.u, y=DDB, uy=DD.u)
  abB[k,] = oNLB$solution[1:2]
}
c(apply(abB, 2, sd), cor(abB)[1,2])

```

A.10 Example 8.8.A

NIST SRM 606 Trace elements in basalt

```

fSkewNormal = function(xoa, mu, sigma, Q025, Q975,
                      mu.weight=1, sigma.weight=1,
                      Q025.weight=1, Q975.weight=1)
{
  xi = xoa[1]; omega = xoa[2]; alpha = xoa[3]
  if (omega < 0) {return(Inf)}
  } else {
    delta = alpha/sqrt(1+alpha^2)
    mu.model = xi + omega*delta*sqrt(2/pi)
    sigma.model = omega*sqrt(1-2*delta^2/pi)
    Q025.model = qsn(0.025, xi=xi, omega=omega, alpha=alpha, solver="RFB")
    Q975.model = qsn(0.975, xi=xi, omega=omega, alpha=alpha, solver="RFB")
    ss = mu.weight*(mu - mu.model)^2
  }
}

```

```

    ss = ss + sigma.weight*(log(sigma / sigma.model))^2
    ss = ss + Q025.weight*(Q025 - Q025.model)^2
    ss = ss + Q975.weight*(Q975 - Q975.model)^2
    return(ss) }
}

skewer = function (x.mu, x.sigma, x.Q025, x.Q975,
                  mu.weight=1, sigma.weight=1, Q025.weight=1, Q975.weight=1)
{
  x.Start = c(x.mu, x.sigma, 2)
  x.o = optim(par=x.Start, fn=fSkewNormal, method="Nelder-Mead",
             mu=x.mu, sigma=x.sigma, Q025=x.Q025, Q975=x.Q975)
  x.xi = x.o$par[1]; x.omega = x.o$par[2]; x.alpha = x.o$par[3]
  x.delta = x.alpha/sqrt(1+x.alpha^2)
  mu.HAT = x.xi + x.omega * x.delta * sqrt(2/pi)
  sigma.HAT = x.omega*sqrt(1-2*x.delta^2/pi)
  Q025.HAT = qsn(0.025, xi=x.xi, omega=x.omega, alpha=x.alpha, solver="RFB")
  Q975.HAT = qsn(0.975, xi=x.xi, omega=x.omega, alpha=x.alpha, solver="RFB")

  curve(dsn(x, xi=x.xi, omega=x.omega, alpha=x.alpha),
        from=mu.HAT-4*sigma.HAT, to=mu.HAT+4*sigma.HAT, n=1001,
        xlab="Measured Value", ylab="Prob. Density",
        col="Blue", lwd=2, axes=FALSE)
  axis(1, lwd=0.5); axis(2, lwd=0.5)
  abline(h=0, col="Gray", lwd=0.5)
  points(c(x.mu, x.Q025, x.Q975), rep(0, 3), pch=18, cex=1, col="Red")
  points(c(mu.HAT, Q025.HAT, Q975.HAT), rep(0, 3),
        pch=5, cex=1.5, col="Red")
  segments(x.mu-x.sigma, 2*par()$cxy[2],
          x.mu+x.sigma, 2*par()$cxy[2], col="Red")
  segments(mu.HAT-sigma.HAT, 2.5*par()$cxy[2],
          mu.HAT+sigma.HAT, 2.5*par()$cxy[2], lty=3, col="Red")

  return(invisible(list(xi=x.xi, omega=x.omega, alpha=x.alpha,
                      mu=mu.HAT, sigma=sigma.HAT,
                      Q025=Q025.HAT, Q975=Q975.HAT)))
}

symmetricalBootstrapCI = function (x, estimate, coverage) {

  ## INPUT : x          = Numeric vector with bootstrap replicates
  ##         estimate = Value where to center coverage interval
  ##         coverage = Coverage probability (between 0 and 1)
  ## OUTPUT: Expanded uncertainty U95 for 95 % coverage

  m = length(x)
  Umax = max(abs(x - estimate))
  Udelta = (max(x)-min(x))/m
  x1 = Umax
  x0 = 0
  iteration = 1
  while ((iteration<1e7) && ((x1-x0)>Udelta))
  {
    iteration = iteration + 1
    x3 = (x1 + x0) / 2
    if(sum((estimate-x3 <= x) &
          (x <= estimate+x3))/m < coverage) {
      x0 = x3 } else { x1 = x3 }
  }
  return((x1 + x0)/2)
}

```

```

Ba = list(w=174, uw=5, Lwr95=166, Upr95=186)
Sr = list(w=169, uw=5, Lwr95=160, Upr95=178)

Ba.skew = skewer(Ba$w, Ba$uw, Ba$Lwr95, Ba$Upr95)
Sr.skew = skewer(Sr$w, Sr$uw, Sr$Lwr95, Sr$Upr95)

K = 1e6
Ba = rsn(K, xi=Ba.skew$xi, omega=Ba.skew$omega, alpha=Ba.skew$alpha, tau=0)
Sr = rsn(K, xi=Sr.skew$xi, omega=Sr.skew$omega, alpha=Sr.skew$alpha, tau=0)

BaSr = Ba/(Ba+Sr)
BaSr.x = huberM(BaSr)$mu
BaSr.u = Qn(BaSr)
BaSr.U95 = symmetricalBootstrapCI(BaSr, estimate=BaSr.x, coverage=0.95)

D = density(BaSr, from=0.47, to=0.545, adjust=1.25, n=2^{11})
plot(c(0.47, 0.545), range(D$y), type="n", axes=FALSE,
      xlab=expression(italic(w)(plain(Ba))/
                     (italic(w)(plain(Ba))+italic(w)(plain(Sr)))),
      ylab="Prob. Density")
axis(1, lwd=0.5); axis(2, lwd=0.5)
iL = which.min(abs(D$x - (BaSr.x-BaSr.U95)))
iU = which.min(abs(D$x - (BaSr.x+BaSr.U95)))
polygon(c(D$x[iL], D$x[iU], D$x[iU:iL]),
        c(0, 0, D$y[iU:iL]), col="Pink", border="Pink")
abline(h=0, lwd=0.5, col="Gray")
segments(D$x[iL], 0, D$x[iU], 0, lwd=1, col="Red")
lines(D$x, D$y, col="Blue", lwd=2)
points(BaSr.x, 0.25*par()$cxy[2], pch=18)

```