MEASUREMENT UNCERTAINTY AND PROBABILITY

A measurement result is incomplete without a statement of its 'uncertainty' or 'margin of error'. But what does that statement actually tell us? This book employs the principles of measurement, probability and statistics to describe what is meant by a '95% interval of measurement uncertainty' and to show how such an interval can be calculated.

The book argues that the concept of a 'target value' is essential if the principles of probability are relevant. It advocates the use of 'extended classical' statistical methods, such as (i) the propagation of higher moments of error distributions, (ii) the evaluation of 'average confidence' intervals, and (iii) the evaluation of 'conditional confidence' intervals. It also describes the use of the Monte Carlo principle for simulating measurements and for constructing procedures that give valid uncertainty statements.

Useful for researchers and graduate students, this book promotes the correct understanding of the classical statistical viewpoint. It also discusses other philosophies, and it employs clear notation and language to avoid the confusion that exists in this controversial field of applied science.

ROBIN WILLINK is a physicist and mathematical statistician who has been employed in Applied Mathematics at Industrial Research Ltd, the parent body of the National Metrology Institute of New Zealand. He is the author of many articles on the subject of measurement uncertainty, and he has made several notable contributions to the statistics literature.

MEASUREMENT UNCERTAINTY AND PROBABILITY

ROBIN WILLINK

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107021938

First published 2013

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Willink, Robin, 1961–
Measurement uncertainty and probability / Robin Willink.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-02193-8
1. Measurement uncertainty (Statistics) 2. Probabilities. I. Title.
QA276.8.W56 2012
519.2–dc23
2012025873

ISBN 978-1-107-02193-8 Hardback

To my friend Gemma, who hasn't got a clue what this book is about.
But that doesn't matter to either of us.

Contents

Acknowledgements — *page* xi
Introduction — xii

Part I Principles

1 Foundational ideas in measurement — 3
1.1 What is measurement? — 3
1.2 True values and target values — 5
1.3 Error and uncertainty — 7
1.4 Identifying the measurand — 10
1.5 The measurand equation — 12
1.6 Success – and the meaning of a ‘95% uncertainty interval’ — 16
1.7 The goal of a measurement procedure — 19

2 Components of error or uncertainty — 21
2.1 A limitation of classical statistics — 21
2.2 Sources of error in measurement — 23
2.3 Categorization by time-scale and by information — 28
2.4 Remarks — 30

3 Foundational ideas in probability and statistics — 32
3.1 Probability and sureness — 33
3.2 Notation and terminology — 39
3.3 Statistical models and probability models — 43
3.4 Inference and confidence — 45
3.5 Two central limit theorems — 52
3.6 The Monte Carlo method and process simulation — 53

4 The randomization of systematic errors — 56

	4.1	The Working Group of 1980	57
	4.2	From classical repetition to practical success rate	59
	4.3	But what success rate? Whose uncertainty?	63
	4.4	Parent distributions for systematic errors	65
	4.5	Chapter summary	70
5	Beyond the ordinary confidence interval		72
	5.1	Practical statistics – the idea of average confidence	72
	5.2	Conditional confidence intervals	77
	5.3	Chapter summary	80
Part II	**Evaluation of uncertainty**		
6	Final preparation		85
	6.1	Restatement of principles	85
	6.2	Two important results	87
	6.3	Writing the measurement model	88
	6.4	Treating a normal error with unknown underlying variance	92
7	Evaluation using the linear approximation		98
	7.1	Linear approximation to the measurand equation	98
	7.2	Evaluation assuming approximate normality	103
	7.3	Evaluation using higher moments	105
	7.4	Monte Carlo evaluation of the error distribution	120
8	Evaluation without the linear approximation		125
	8.1	Including higher-order terms	125
	8.2	The influence of fixed estimates	130
	8.3	Monte Carlo simulation of the measurement	135
	8.4	Monte Carlo simulation – an improved procedure	141
	8.5	Monte Carlo simulation – variant types of interval	145
	8.6	Summarizing comments	150
9	Uncertainty information fit for purpose		152
	9.1	Information for error propagation	152
	9.2	An alternative propagation equation	157
	9.3	Separate variances – measurement of functionals	161
	9.4	Worst-case errors – measurement in product testing	164
	9.5	Time-scales for errors	169
Part III	**Related topics**		
10	Measurement of vectors and functions		175
	10.1	Confidence regions	176

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

10.2 Simultaneous confidence intervals 179
10.3 Data snooping and cherry picking 184
10.4 Curve fitting and function estimation 185
10.5 Calibration 189

11 Why take part in a measurement comparison? 192
11.1 Examining the uncertainty statement 192
11.2 Estimation of the Type B error 198
11.3 Experimental bias 201
11.4 Complicating factors 203

12 Other philosophies 204
12.1 Worst-case errors 204
12.2 Fiducial inference 205
12.3 Bayesian inference 207
12.4 Measurement without target values 214

13 An assessment of objective Bayesian statistics 220
13.1 Ignorance and coherence 222
13.2 Information and entropy 230
13.3 Meaning and communication 234
13.4 A presupposition? 235
13.5 Discussion 236

14 *Guide to the Expression of Uncertainty in Measurement* 237
14.1 Report and recommendation 237
14.2 The mixing of philosophies 239
14.3 Consistency with other fields of science 241
14.4 Righting the *Guide* 242

15 Measurement near a limit – an insoluble problem? 245
15.1 Formulation 245
15.2 The Feldman–Cousins solution 247
15.3 The objective Bayesian solution 250
15.4 Purpose and realism 253
15.5 Doubts about uncertainty 255
15.6 Conclusion 257

Appendix A The weak law of large numbers 259
Appendix B The Sleeping Beauty paradox 260
Appendix C The sum of normal and uniform variates 261
Appendix D Analysis with one Type A and one Type B error 262
Appendix E Conservatism of treatment of Type A errors 263

Appendix F An alternative to a symmetric beta distribution 264
Appendix G Dimensions of the ellipsoidal confidence region 266
Appendix H Derivation of the Feldman–Cousins interval 267

References 268
Index 273

Various quotations at the beginnings of Parts I, II and III were obtained from "Statistically speaking: a dictionary of quotations", selected and arranged by C. C. Gaither and A. E. Cavazos-Gaither. The last is due to Paul Gardner, writer-producer of "Richard Tuttle: Never Not an Artist".

The author can be contacted at robin.willink@gmail.com

Acknowledgements

This book is borne out of a respect for the ideas of truth and meaning. That respect is associated with my trust in God, whom I see as having given me the motivation, ability and courage required. The material contained has been informed by many past and present discussions with scientists at the Measurement Standards Laboratory of New Zealand; I am especially grateful for the support of Blair Hall, Jeremy Lovell-Smith, Rod White and Annette Koo. I would also like to thank an anonymous referee of my proposal to Cambridge University Press about this book. In submitting a report, the referee wrote to the publisher 'You offered me to choose a Cambridge book: if it is allowed to me, I will choose exactly a copy of this book!' What a wonderful, affirming, remark.

Introduction

I have recently finished a period of employment as a statistician at a research institute that, in one capacity, acts as the parent body for New Zealand's national laboratory for metrology (measurement science). During that period, a significant amount of time was spent considering and discussing ideas of data-analysis in physics and chemistry, especially in matters relating to measurement. I quickly reached the conclusion that the ideas surrounding the quantification of measurement uncertainty are problematic, and later I formed some opinions as to why that is the case. First, the evaluation of uncertainty is supposed to have been adequately addressed in the *Guide to the Expression of Uncertainty in Measurement* (BIPM *et al.*, 1995; JCGM, 2008a) but – as is described in Chapter 14 – that document is unsatisfactory from a statistical point of view. Second, for typical measurement scientists, the evaluation of uncertainty in measurement is not the most interesting aspect of their work. Less effort is expended in understanding conceptual matters of uncertainty than in addressing practical matters of measurement technique. Third, and perhaps most important, the task involves concepts of statistical inference that might not have been presented well during an education in a science faculty.

The statistician involved in this area of science might reasonably expect to find one outstanding difficulty: the 'Bayesian controversy' that splits the statistical community over the nature of 'probability'. This controversy is surely relevant in measurement problems, where systematic errors are ubiquitous and where these errors might be treated by imputing worst-case values – as a classical statistician often would – or be treated 'probabilistically' – as a Bayesian statistician would. But the statistician might be surprised to also find another difficulty: the measurement community seems divided over whether quantities can be considered to have 'true' values. That issue is of great relevance. The ground is cut out from underneath someone carrying out parameter estimation if a 'true value' for the parameter does not exist. Yet the concepts of such parameter estimation are those on which measurement scientists appear to base their treatments of measurement uncertainty. Thus, there is this curiosity: some measurement scientists are implicitly denying the very basis of the methods of data analysis that they employ.

So the statistician seeking to make sense of what metrologists understand by 'the quantification of measurement uncertainty' has a surprisingly difficult task. This was emphasized to me when, in the library at my place of work, I viewed a copy of the book *Measurement Errors: Theory and Practice* by the experienced and authoritative Russian scientist Semyon Rabinovich (1995). At three places in the margin of the very first page, where Rabonivich introduces basic ideas of measurement, some dissenting reader had written the word 'wrong'! No further evidence was needed to convince me that different measurement scientists think about measurement in fundamentally different ways. If such scientists cannot agree on foundational ideas about measurement then how can the statistician contribute easily in the area of 'measurement uncertainty'? Perhaps some of the uncertainty in measurement is about the concept of measurement itself!

Despite these difficulties – or perhaps because of them – my work at that institute led to the publication of a number of articles of a statistical nature in international measurement journals. The kind words and compliments that I received in response to those articles indicated that many measurement scientists are searching for a coherent understanding of measurement uncertainty based on accepted logical principles. A time came when a colleague in Slovakia asked me 'Robin, have you thought of writing a book on this subject?' I had, but I had not made or taken an opportunity. The opportunity has since presented itself, and this book is the result.

The purpose of the book

I am a physicist-turned-statistician, not an expert in measurement. So this book is not intended to describe subtle aspects of measurement technique. Rather it is intended to explain the various concepts of 'probability' that are applicable in measurement, especially in the production of a justifiable statement of measurement uncertainty. Therefore, where aspects of measurement are described, the emphasis is on ideas that permit the proper use of probability and its ally 'statistics'. There is little more distressing to a scientist than to have the principles of one's discipline misused. So this book is concerned with the correct use of concepts of probability and statistics, and on the legitimate extension of these concepts to address the realities of measurement.

The level at which this book is written is above that of an introductory text. Introductory texts and courses on statistics do not address some of the issues that arise in measurement and often fail to discuss potential areas of confusion for scientists of other disciplines. A simple presentation of basic statistical ideas does not suffice, and for some reason a gulf remains between the understanding of the statistician and that of the typical applied scientist or mathematician. Consequently, this book is written at a higher level, the concepts presented are focused on measurement, and most of the ideas found in introductory statistics books are omitted or assumed. In particular, the emphasis is on reexpressing the relevant principles of data analysis for 'parameter estimation' and on extending these principles to fit the context of measurement.

The physicist, chemist or measurement scientist who reads this book is asked to put aside preconceptions about statistical ideas and terms, especially those reinforced through reading material not written by statisticians. In turn, the statistician who reads this book is asked to engage with the issues that face experimental scientists and to subsequently present the science of statistics in a way that narrows the gulf of understanding (or mis-understanding). Both groups of reader are asked to acknowledge that the language and notation involved must be sufficiently complex to avoid ambiguity. Most importantly, all readers are asked to consider what the actual meaning of an interval of measurement uncertainty should be, especially when this interval is quoted with a figure of 'sureness', say 95%.

That is the central issue raised in this book. It is an issue or question that can be expressed in different forms.

- How can the person for whom a measurement is made properly interpret a 95% interval of measurement uncertainty provided with the result?
- What is the practical implication of the figure 95% in this context?

Readers who are not interested in addressing this question need proceed no further, for the great majority of this book is written to answer it and to develop corresponding means of uncertainty evaluation.

At the forefront of this book is the idea that the meaning of a stated scientific result should be unambiguous. Of course, approximations are unavoidable in modelling and data analysis. But these should be *approximations in number, not approximations in meaning*. It is legitimate to say 'the correct numerical answer is approximately ...' but it does not seem legitimate to say 'the meaning of this answer is approximately ...'. So in this book we require a 95% interval of measurement uncertainty to have an unambiguous and practical meaning. This involves identifying a concept of success in measurement and, by implication, a concept of failure. Without such ideas we have very little to hang our hats upon, very little to claim that we have actually achieved in a measurement, and very little to which we are accountable. Such concepts of achievement and responsibility are not always evident in approaches to the evaluation of uncertainty in measurement.

Structure and style

I trust that this book will be of some benefit to all who engage with it. The book is written in three parts that will be of different value to different readers. Part I describes and explains various principles that provide a basis for the study of measurement uncertainty. It might be lengthier than many readers would expect, but the lack of a set of accepted starting principles is one reason for the confusion and disagreement that we find when approaching this subject. Chapters 1 and 2 consider relevant principles of measurement, in particular the ideas of measurement uncertainty and measurement error. Chapter 3 outlines relevant principles of probability and statistics, explains why the classical approach to statistical analysis is favoured in this book and pays attention to the correct understanding of a *confidence interval*, which is the basic tool of a classical statistician when estimating an unknown parameter using an interval. Chapter 4 considers the idea of treating systematic errors as if they had been drawn from probability distributions, which was an idea recommended by an authoritative group of measurement scientists convened to consider the difficulties of treating these errors. Last, Chapter 5 introduces variant forms of confidence interval that seem particularly applicable in our context of measurement. Many of the ideas developed in Chapters 4 and 5 fall outside the body of classical statistical principles. So the approach taken to the evaluation of measurement uncertainty in this book might be described as one of 'extended classical statistics'.

Part II builds on the principles described in Part I to present methods for the evaluation of uncertainty that are intended to be accurate, meaningful and practical. Chapter 6 completes our preparation for this task by addressing some remaining issues. In particular, it describes a pragmatic step taken to unify the treatment of 'statistical' and 'non-statistical' errors. Chapter 7 discusses the evaluation of measurement uncertainty based on the first-order Taylor's series approximation to the equation defining the 'measurand', the quantity intended to be measured. This chapter describes methodology sufficient for the evaluation of uncertainty in the vast majority of measurement problems. In particular, it puts forward a method that involves the first four moments of the total error distribution, not just the mean and variance. Chapter 8 considers the evaluation of measurement uncertainty when the full non-linear function is retained. In particular, it discusses the use of Monte Carlo simulation of the measurement process. This chapter involves the most complicated mathematical ideas found in this book, and much of it may be omitted. Chapter 9 encourages us to take into account the purpose of making the measurement when considering the appropriate way of representing the measurement uncertainty.

Part III addresses miscellaneous topics. Chapter 10 discusses the measurement of vectors and functions, and Chapter 11 considers the analysis of data from an inter-laboratory comparison conducted to assess, and perhaps refine, a statement of measurement uncertainty. These two chapters largely stand alone, so they may be omitted in a first reading. The later chapters seek to explain and overcome sources of confusion and debate. Readers are asked to study these chapters carefully: the comments contained might be the most helpful found in this book! Chapter 12 considers other approaches to the evaluation of measurement uncertainty, while Chapter 13 focuses on the 'objective Bayesian' approach, which is being advocated in some official documents. Chapter 14 discusses the principles of the *Guide to the Expression of Uncertainty in Measurement* (BIPM *et al.*, 1995; JCGM, 2008a), which is widely seen as an international standard. Last, Chapter 15 discusses the difficult situation where the value of the measurand lies close to a non-physical region into which the interval of measurement uncertainty cannot extend. This situation highlights some of the philosophical issues that motivate the writing of this book.

The text of the book contains sections, examples, definitions and results that are numbered. There are pieces of theory that are numbered and there are also three numbered 'claims'. These items have been written to stand alone. They are numbered within each chapter, and each type of item has its own counter. So Example 3.2 is the second example found in Chapter 3, it is not necessarily found in Section 3.2, and it might precede Definition 3.1.

I have tried to write in a style that is consistent, clear and interesting, and I hope that reading this book will be both informative and enjoyable. I like to think that the book contains new ideas, and that the reader might be surprised by various assertions and so be encouraged to take fresh looks at old problems. Such assertions will be largely associated with two underlying concepts emphasized throughout, these being 'meaning' and 'purpose'.

I have also tried to convey something of my enthusiasm for the classical approach to statistical inference, while not being unduly critical of other points of view. In one or two places – perhaps late in Chapter 8 – this book might be seen as advocating theory that is inapplicable through being impractical. However, some other approaches seem inapplicable through being meaningless or invalid! (It does not seem possible to have an interest in probability and statistics while not having a view of what is meaningful.)

Some simple terms

One theme of this book relates to clarity in notation and terminology. The usage of notation is discussed in some length in Section 3.2, while matters of language and terminology are raised throughout the text. The existence of unambiguous language in science is very important, especially in an interdisciplinary field like this – where ideas of measurement, physics and statistics must meet if progress is to be made. So let me state at the outset the intended meanings of some seemingly simple words. I am convinced that many of the problems we face in this subject arise from different understandings of such words.

- The words *quantity* and *value* will be used in very general senses rather than in any sense that is peculiar to the measurement scientist.
- The word *fixed* will be used to refer to something that cannot change during a process, even though it may be unknown. So a fixed quantity might be regarded as having been set by Providence rather than determined by humankind.
- The verb *to estimate* will mean 'to approximate by some reasonable means'. For something to be approximated, that thing must exist. So for something to be estimated, that thing must exist.[1]
- The verb *to determine* will not be used often. It conveys a much stronger idea than the verb 'to estimate'. To determine the value of something is to find the value exactly. Some scientists speak of 'determining the value of some physical quantity', but I do not understand what they mean. Are they claiming to decree what the value of the quantity is – as in one of the views of measurement discussed in Chapter 1 – or are they using the word 'determine' as others would use the word 'estimate'?
- The word *bias* will be used in a way that conforms to its use in statistics, where the bias is *the error that would be incurred if the estimation procedure were repeated infinitely and the individual results were averaged to form the final result.* This understanding contrasts with the definition of 'measurement bias' given in the *International Vocabulary of Metrology* (VIM, 2012, clause 2.18), which is 'estimate of a **systematic measurement error**', the term in bold being defined in the preceding clause. That definition seems to be an aberration: most scientists would be much more comfortable with the idea that a bias is something to be estimated, not an estimate of something.

This book is written to address the controversies and divisions that exist within this field, some of which are exacerbated by ambiguities about such words. I trust that the book will shed some light on areas in which scientists disagree, especially where there are informal self-taught views of probability and measurement uncertainty. And I would like it to enable the measurement scientist to employ methods acceptable to those concerned about the principles of probability and statistics. In addition, I hope that statisticians reading this book might see the issues that face the measurement scientist and might realize that developing a suitable theory of measurement uncertainty requires thinking outside the statistical square.

[1] The claim that we can 'estimate the values of model parameters' when the model and its parameters are known to be abstractions of reality seems to me to be a corruption of English language that has led to much illogical practice.