Bayesian Reasoning in Data Analysis

A Critical Introduction

This page is intentionally left blank

Bayesian Reasoning in Data Analysis

A Critical Introduction

Giulio D'Agostini

*Physics Department
University of Rome "La Sapienza", Italy*

World Scientific

New Jersey • London • Singapore • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

D'Agostini, G. (Giulio).

Bayesian reasoning in data analysis : a critical introduction / Giulio

D'Agostini.

p. cm.

Includes bibliographical references and index.

ISBN 981-238-356-5 (alk. paper)

1. Bayesian statistical decision theory. I. Title.

QA279.5 .D28 2003

2003045082

519.5'42--dc21

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2003 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Printed in Singapore by World Scientific Printers (S) Pte Ltd

To my parents, who always told me
“credi quello che ti puoi credere”
(*“believe what you find it reasonable to believe”*).

This page is intentionally left blank

Preface

This book is primarily addressed to physicists and other scientists and engineers who need to evaluate uncertainty in measurement. However, a large portion of its contents should be of interest to anyone who has to deal with probability and uncertainty, has an elementary background in ‘standard statistics’ and is wondering what this ‘new statistics reasoning’ is all about.

Although, like myself, you might never have heard about Bayes’ theorem during your studies, in recent years you will almost certainly have encountered, with increasing frequency, the term ‘Bayesian’ in articles, books and the media. The so-called Bayesian methods are being employed in the most disparate fields of research and application, from engineering to computer science, economics, medicine and even forensic science. Some people are going so far as to talk of a ‘paradigm shift’, in the Kuhnian sense, although it is a strange revolution indeed which has its roots in the centuries-old ideas of the founding fathers of probability - the likes of Bernoulli, Bayes, Laplace and Gauss.

The gist of Bayesian statistics is not difficult to grasp. At its base is the intuitive idea that probability quantifies the ‘degree of belief’ in an event (in this context, an *event* is whatever can be precisely described by a proposition). Now, degrees of belief can be framed in a mathematical structure which allows the probability of an event A to be calculated on the basis of the probability of other events logically connected to that event A . In particular, the probability of event A changes if other events are assumed to be ‘true’, provided these other events are ‘stochastically dependent’ on event A . This is the essence of Bayes’ theorem. As a consequence, Bayesian statistics allows the probability of a hypothesis-event to be continually updated on the basis of new observation-events that depend on that hypothesis-event.

Most likely this is not the way you were taught (elements of) probability theory. At most you might have been warned about the existence of a 'subjective probability' as an 'interpretation of probability', helpful in many fields, but definitively not applicable in Science where 'you want to be objective'. This is what I was taught in my training and was the approach I adopted in research, until I suddenly realized that there was something wrong with those ideas and with the methods which resulted from them. The breakthrough came when I myself had to teach probability and data analysis. Here is how several years later I reported my personal experience in the bulletin of the International Society for Bayesian Analysis (ISBA Newsletter, March 2000).

It is well known that the best way to learn something is to teach it. When I had to give the Laboratory of Physics course to Chemistry students and introduce elements of probability and statistics applied to data analysis, I did as most new teachers do: I started repeating what I had learned years before, more or less using the same lecture notes. This worked well for explaining the experiments, but when I moved to probability, the situation was quite embarrassing. In the very first lecture I realized that I was not convinced of what I was saying. I introduced probability as the ratio between favorable and possible cases, but I had no courage to add 'if the cases are equally probable'. I cheated by saying 'if the cases are equally possible' and moved rapidly to examples. The students had no time to react, the examples were well chosen, and I was able to survive that lesson and the following weeks.

The problem returned when we came to the evaluation of measurement uncertainty, a typical application of statistics in scientific disciplines. I had to acknowledge that the reasoning physicists actually use in practice is quite in contradiction with the statistics theory we learn and teach. The result was that, whereas I had started the semester saying that subjective probability was not scientific, I ended it teaching probability inversion applied to physics quantities.

I cannot speak of a 'conversion' to Bayesianism, because at that time (spring 1993) I had no alternative framework at my disposal. All books and lecture notes I had were strictly 'standard'. Just one book for economics students, which arrived more or less by chance on my desk, contained, as a kind of side remark, some examples of applications of Bayes' theorem. The problems were so trivial that anyone with a bit of imagination could have

solved them just by building contingency tables, without needing to resort to probability theory and that ‘strange theorem’. Subsequently, I worked out more intriguing examples, I extended Bayes’ theorem to continuous variables and applied it to typical measurements in physics. Only months later did I discover that a Bayesian community existed, that my results had been known for two centuries and that there were (and there *are*) heated debates between ‘Bayesians’ and ‘frequentists’. I was a bit disappointed to learn that my wheel had already been invented centuries earlier (and it was turning so fast I could barely keep up with it!). But I was also glad to realize that I was in good company. Only at this stage did I start to read the literature and to clarify my ideas.

I consider this initial self-learning process to have been very important because, instead of being ‘indoctrinated’ by a teacher, as had happened with my frequentistic training, I was instinctively selecting what was more in tune with my intuitive ideas and my fifteen years of experience in frontier physics.

I am sorry to be bothering you with these autobiographical notes, but I think they will help in understanding the spirit in which this book has been written and its idiosyncratic style. (I hope, at least, you will find it less tedious than the average statistics book.) For instance, you will find continual, sometimes sarcastic, criticisms of ‘conventional’ statistical methods. I even refuse to call these methods ‘classical’, because this appellation is misleading too (the term ‘classical’ usually refers to the approach of the founding fathers of a subject area, but the reasoning of the pioneers of probability theory was closer to what we nowadays call Bayesian). You might wonder why I am so doggedly critical of these conventional methods. The reason is that I feel I have been cheated by names and methods which seem to mean something they do not. I therefore make it a central issue in this book to show, by reasoning and examples, why many standard statistical recipes are basically wrong, even if they can often produce reasonable results. I simply apply scientific methodology to statistical reasoning in the same way as we apply it in Physics and in Science in general. If, for example, experiments show that Parity is violated, we can be disappointed, but we simply give up the principle of Parity Conservation, at least in the kind of interactions in which it has been observed that it does not hold. I do not understand why most of my colleagues do not behave in a similar way with the Maximum Likelihood principle, or with the ‘prescriptions’ for building Confidence Intervals, both of which are known to produce absurd results. At most, these methods should be used for special well-controlled

cases, under well-stated assumptions.

To continue with my story, some months after having taught my first course, in an attempt to invent interesting problems for students, I wrote a little program for spectrum deconvolution ('unfolding' in our jargon) which several particle and astro-particle physics teams subsequently found useful for analyzing their data. This gave me more confidence in Bayesian ideas (but I continued to avoid the adjective 'subjective', which I still considered negative at that time) and I began to give seminars and mini-courses on the subject. In particular, lectures I gave in 1995 to graduate students at University of Rome 'La Sapienza' and to summer students at the Deutsches Elektronen-Synchrotron in Hamburg, Germany, encouraged me to write the 'Bayesian Primer' (DESY-95-242, Roma1 N. 1070), which forms the core of this book. I took advantage of the 'academic training' course I gave to researchers of the European Organization for Nuclear Physics in 1998 to add some material and turn the Primer into CERN Report 99-03. The final step towards producing this book was taken in 2002, thanks to the interest of World Scientific in publishing an expanded version of the previous reports.

Instead of completely rewriting the Primer, producing a thicker report which would have been harder to read sequentially, I have divided the text into three Parts.

- Part 1 is devoted to a critical review of standard statistical methods and to a general overview of the proposed alternative. It contains references to the other two Parts for details.
- Part 2 is an extension of the original Primer, subdivided into chapters for easier reading.
- Part 3 contains further comments concerning the general aspects of probability, as well as other applications.

The advantage of this structure is that the reader should be able to get an overall view of problems and proposed solutions and then decide if he or she wants to enter into details. I hope this organization of the contents will suit the typical reader, whom I find it hard to imagine wishing to read sequentially a tome of over three hundred pages! This structure also allows the book to be read at several levels. For example, most of chapters 1, 2, 3, 5 and 10, which are the most important as far as the basic ideas are concerned, do not require advanced mathematical skills and can be understood by the general reader. However, organizing things in this manner has inevitably led to some repetition. I have tried to keep repetitions to a minimum, but *repetita juvant*, especially in this subject where the real difficulty lies not

in understanding the formalism, but in shaking off deep-rooted prejudices.

A comment about the title of this book is in order. A title closer to the spirit of the approach proposed here would have been “Probabilistic reasoning . . .”. In fact, the term ‘Bayesian’ might seem somewhat narrow, as if I am implying that the methods illustrated here always require explicit use of Bayes’ theorem. However, in common usage, ‘Bayesian’ has come to mean ‘based on the intuitive idea of probability’. Thus, what is known as the Bayesian approach is effectively a theory of uncertainty which is applicable universally. Within it, ‘probability’ has the same meaning for everybody: precisely that meaning which the human mind has developed naturally and which frequentists have tried to kill. Therefore, I have kept the term ‘Bayesian’ in the title, with the hope of attracting the attention of those who are curious about what ‘Bayesian’ might mean.

This book is based on the work of several years, during which I have had the opportunity to interact, directly or indirectly, with a large variety of persons, most of them physicists and physics students of many nationalities, but also mathematicians, statisticians, metrologists and science historians and philosophers. In particular, the interest shown by those who attended the lectures, and also the criticisms of those who had strong prejudices towards the approach I was presenting, has been highly stimulating. I take this opportunity to thank them all. Special acknowledgements go to Romano Scozzafava for many discussions about the fundamental aspects of probability theory. The many clarifications about DIN and ISO recommendations received by Klaus Weise of the PTB Braunschweig (Germany) have been particularly useful. I would like to thank Paolo Agnoli, Pia Astone, Peppe Degrassi, Volker Dose, Fritz Fröhner, Ken Hanson, Frank Lad, Daniela Monaldi, Gianni Penso, Mirko Raso, Stefan Schlenstedt, Myron Tribus and Günter Zech for discussions and critical comments on the manuscript, as well on the on the old version of the Primer [1] and the CERN Report [2] on which this book is based. Finally I would like to thank Jim Mc Manus for his help in finding ways to better express my ideas in English (apart from this sentence) and Bruno Pellizzoni for technical support with many of the drawings.

Rome, March 2003

Email: giulio.dagostini@roma1.infn.it

URL: <http://www.roma1.infn.it/~dagos/>

This page is intentionally left blank

Contents

Part 1 Critical review and outline of the Bayesian alternative	1
1. Uncertainty in physics and the usual methods of handling it	3
1.1 Uncertainty in physics	3
1.2 True value, error and uncertainty	5
1.3 Sources of measurement uncertainty	6
1.4 Usual handling of measurement uncertainties	7
1.5 Probability of observables versus probability of ‘true values’	9
1.6 Probability of the causes	11
1.7 Unsuitability of frequentistic confidence intervals	11
1.8 Misunderstandings caused by the standard paradigm of hypothesis tests	15
1.9 Statistical significance versus probability of hypotheses .	19
2. A probabilistic theory of measurement uncertainty	25
2.1 Where to restart from?	25
2.2 Concepts of probability	27
2.3 Subjective probability	29
2.4 Learning from observations: the ‘problem of induction’ .	32
2.5 Beyond Popper’s falsification scheme	34

2.6	From the probability of the effects to the probability of the causes	34
2.7	Bayes' theorem for uncertain quantities	36
2.8	Afraid of 'prejudices'? Logical necessity versus frequent practical irrelevance of priors	37
2.9	Recovering standard methods and short-cuts to Bayesian reasoning	39
2.10	Evaluation of measurement uncertainty: general scheme .	41
2.10.1	Direct measurement in the absence of systematic errors	41
2.10.2	Indirect measurements	42
2.10.3	Systematic errors	43
2.10.4	Approximate methods	46

Part 2 A Bayesian primer 49

3.	Subjective probability and Bayes' theorem	51
3.1	What is probability?	51
3.2	Subjective definition of probability	52
3.3	Rules of probability	55
3.4	Subjective probability and 'objective' description of the physical world	58
3.5	Conditional probability and Bayes' theorem	60
3.5.1	Dependence of the probability on the state of information	60
3.5.2	Conditional probability	61
3.5.3	Bayes' theorem	63
3.5.4	'Conventional' use of Bayes' theorem	66
3.6	Bayesian statistics: learning by experience	68
3.7	Hypothesis 'test' (discrete case)	71
3.7.1	Variations over a problem to Newton	72
3.8	Falsificationism and Bayesian statistics	76
3.9	Probability versus decision	76
3.10	Probability of hypotheses versus probability of observations	77
3.11	Choice of the initial probabilities (discrete case)	78
3.11.1	General criteria	78
3.11.2	Insufficient reason and Maximum Entropy	81

3.12	Solution to some problems	82
3.12.1	AIDS test	82
3.12.2	Gold/silver ring problem	83
3.12.3	Regular or double-head coin?	84
3.12.4	Which random generator is responsible for the observed number?	84
3.13	Some further examples showing the crucial role of background knowledge	85
4.	Probability distributions (a concise reminder)	89
4.1	Discrete variables	89
4.2	Continuous variables: probability and probability density function	92
4.3	Distribution of several random variables	98
4.4	Propagation of uncertainty	104
4.5	Central limit theorem	108
4.5.1	Terms and role	108
4.5.2	Distribution of a sample average	111
4.5.3	Normal approximation of the binomial and of the Poisson distribution	111
4.5.4	Normal distribution of measurement errors . . .	112
4.5.5	Caution	112
4.6	Laws of large numbers	113
5.	Bayesian inference of continuous quantities	115
5.1	Measurement error and measurement uncertainty	115
5.1.1	General form of Bayesian inference	116
5.2	Bayesian inference and maximum likelihood	118
5.3	The dog, the hunter and the biased Bayesian estimators .	119
5.4	Choice of the initial probability density function	120
5.4.1	Difference with respect to the discrete case . . .	120
5.4.2	Bertrand paradox and angels' sex	121
6.	Gaussian likelihood	123
6.1	Normally distributed observables	123
6.2	Final distribution, prevision and credibility intervals of the true value	124
6.3	Combination of several measurements – Role of priors . .	125

6.3.1	Update of estimates in terms of Kalman filter . .	126
6.4	Conjugate priors	126
6.5	Improper priors — never take models literally!	127
6.6	Predictive distribution	127
6.7	Measurements close to the edge of the physical region . .	128
6.8	Uncertainty of the instrument scale offset	131
6.9	Correction for known systematic errors	133
6.10	Measuring two quantities with the same instrument having an uncertainty of the scale offset	133
6.11	Indirect calibration	136
6.12	The Gauss derivation of the Gaussian	137
7.	Counting experiments	141
7.1	Binomially distributed observables	141
7.1.1	Observing 0% or 100%	145
7.1.2	Combination of independent measurements . . .	146
7.1.3	Conjugate prior and many data limit	146
7.2	The Bayes problem	148
7.3	Predicting relative frequencies — Terms and interpretation of Bernoulli's theorem	148
7.4	Poisson distributed observables	152
7.4.1	Observation of zero counts	154
7.5	Conjugate prior of the Poisson likelihood	155
7.6	Predicting future counts	155
7.7	A deeper look to the Poissonian case	156
7.7.1	Dependence on priors — practical examples . . .	156
7.7.2	Combination of results from similar experiments	158
7.7.3	Combination of results: general case	160
7.7.4	Including systematic effects	162
7.7.5	Counting measurements in the presence of background	165
8.	Bypassing Bayes' theorem for routine applications	169
8.1	Maximum likelihood and least squares as particular cases of Bayesian inference	169
8.2	Linear fit	172
8.3	Linear fit with errors on both axes	175
8.4	More complex cases	176

8.5	Systematic errors and ‘integrated likelihood’	177
8.6	Linearization of the effects of influence quantities and approximate formulae	178
8.7	BIPM and ISO recommendations	181
8.8	Evaluation of type B uncertainties	183
8.9	Examples of type B uncertainties	184
8.10	Comments on the use of type B uncertainties	186
8.11	Caveat concerning the blind use of approximate methods	189
8.12	Propagation of uncertainty	191
8.13	Covariance matrix of experimental results – more details	192
8.13.1	Building the covariance matrix of experimental data	192
8.13.1.1	Offset uncertainty	193
8.13.1.2	Normalization uncertainty	195
8.13.1.3	General case	196
8.14	Use and misuse of the covariance matrix to fit correlated data	197
8.14.1	Best estimate of the true value from two correlated values	197
8.14.2	Offset uncertainty	198
8.14.3	Normalization uncertainty	198
8.14.4	Peelle’s Pertinent Puzzle	202
9.	Bayesian unfolding	203
9.1	Problem and typical solutions	203
9.2	Bayes’ theorem stated in terms of causes and effects	204
9.3	Unfolding an experimental distribution	205

Part 3 Further comments, examples and applications 209

10.	Miscellanea on general issues in probability and inference	211
10.1	Unifying role of subjective approach	211
10.2	Frequentists and combinatorial evaluation of probability	213
10.3	Interpretation of conditional probability	215
10.4	Are the beliefs in contradiction to the perceived objectivity of physics?	216
10.5	Frequentists and Bayesian ‘sects’	220

10.5.1	Bayesian versus frequentistic methods	221
10.5.2	Subjective or objective Bayesian theory?	222
10.5.3	Bayes' theorem is not everything	226
10.6	Biased Bayesian estimators and Monte Carlo checks of Bayesian procedures	226
10.7	Frequentistic coverage	229
10.7.1	Orthodox teacher versus sharp student - a dialogue by George Gabor	232
10.8	Why do frequentistic hypothesis tests 'often work'? . . .	233
10.9	Comparing 'complex' hypotheses – automatic Ockham' Razor	239
10.10	Bayesian networks	241
10.10.1	Networks of beliefs – conceptual and practical applications	241
10.10.2	The gold/silver ring problem in terms of Bayesian networks	242
11.	Combination of experimental results: a closer look	247
11.1	Use and misuse of the standard combination rule	247
11.2	'Apparently incompatible' experimental results	249
11.3	Sceptical combination of experimental results	252
11.3.1	Application to ϵ'/ϵ	259
11.3.2	Posterior evaluation of σ_i	262
12.	Asymmetric uncertainties and nonlinear propagation	267
12.1	Usual combination of 'statistic and systematic errors' . . .	267
12.2	Sources of asymmetric uncertainties in standard statistical procedures	269
12.2.1	Asymmetric χ^2 and ' $\Delta\chi^2 = 1$ rule'	269
12.2.2	Systematic effects	272
12.2.2.1	Asymmetric beliefs on systematic effects	273
12.2.2.2	Nonlinear propagation of uncertainties	273
12.3	General solution of the problem	273
12.4	Approximate solution	275
12.4.1	Linear expansion around $E(\mathbf{X})$	276
12.4.2	Small deviations from linearity	278

12.5	Numerical examples	280
12.6	The non-monotonic case	282
13.	Which priors for frontier physics?	285
13.1	Frontier physics measurements at the limit to the detector sensitivity	285
13.2	Desiderata for an optimal report of search results	286
13.3	Master example: Inferring the intensity of a Poisson process in the presence of background	287
13.4	Modelling the inferential process	288
13.5	Choice of priors	288
13.5.1	Uniform prior	289
13.5.2	Jeffreys' prior	290
13.5.3	Role of priors	292
13.5.4	Priors reflecting the positive attitude of researchers	292
13.6	Prior-free presentation of the experimental evidence	295
13.7	Some examples of $\mathcal{R}$ -function based on real data	298
13.8	Sensitivity bound versus probabilistic bound	299
13.9	Open versus closed likelihood	302

Part 4 Conclusion 305

14.	Conclusions and bibliography	307
14.1	About subjective probability and Bayesian inference	307
14.2	Conservative or realistic uncertainty evaluation?	308
14.3	Assessment of uncertainty is not a mathematical game	310
14.4	Bibliographic note	310

<i>Bibliography</i>	313
-------------------------------	-----

<i>Index</i>	325
------------------------	-----