Data Analysis Techniques for Physical Scientists

Data Analysis Techniques for Physical Scientists is a comprehensive guide to data analysis techniques for physical scientists, providing a valuable resource for advanced undergraduate and graduate students, as well as seasoned researchers. The book begins with an extensive discussion of the foundational concepts and methods of probability and statistics under both the frequentist and Bayesian interpretations of probability. It next presents basic concepts and techniques used for measurements of particle production cross sections, correlation functions, and particle identification. Much attention is devoted to notions of statistical and systematic errors, beginning with intuitive discussions and progressively introducing the more formal concepts of confidence intervals, credible range, and hypothesis testing. The book also includes an in-depth discussion of the methods used to unfold or correct data for instrumental effects associated with measurement and process noise as well as particle and event losses, before ending with a presentation of elementary Monte Carlo techniques. This title is also available as open access on Cambridge Core.

Claude A. Pruneau is a Professor of Physics at Wayne State University, from where he received the 2006 Excellence in Teaching Presidential award. He is also a member of the ALICE collaboration, and conducts an active research program in the study of the Quark Gluon Plasma produced in relativistic heavy ion collisions at the CERN Large Hadron Collider. He has worked as a Research Fellow at both Atomic Energy for Canada Limited and McGill University, and is a member of the American Physical Society, Canadian Association of Physicists and the Union of Concerned Scientists.

Data Analysis Techniques for Physical Scientists

CLAUDE A. PRUNEAU
Wayne State University, Michigan

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

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi - 110002, India

79 Anson Road, #06-04/06, Singapore 079906

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/9781108416788
DOI: 10.1017/9781108241922

When citing this work, please include a reference to the DOI 10.1017/9781108241922

First published 2017

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

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

ISBN 978-1-108-41678-8 Hardback
ISBN 978-1-009-24500-5 Paperback

To my son Blake

Funding from the SCOAP3 for Books initiative made it possible for this book to be published open access in 2025, making the digital version freely available for anyone to read under a Creative Commons licence.

Contents

Preface

Physics students typically take a wide range of advanced classes in mechanics, electromagnetism, quantum mechanics, thermodynamics, and statistical mechanics, but sadly, receive only limited formal training in data analysis techniques. Most students in experimental physics indeed end up gleaning the required material by reading parts of a plurality of books and scientific articles. They typically end up knowing a lot about one particular analysis technique but relatively little about others. Paradoxically, modern experiments in particle and nuclear physics enable an amazingly wide range of very sophisticated measurements based on diverse analytical techniques. The end result is that beginning students may have a rather limited understanding of the many papers they become coauthors of by virtue of being members of a large scientific collaboration. After twenty years of teaching "physics" and carrying out research in heavy-ion physics, I figured I should make an effort to remedy this situation by creating a book that covers all the basic tools required in the data analysis of experiments at RHIC, the LHC, and other large experimental facilities.

This was a fairly ambitious project given that the range of techniques employed in today's experiments is actually quite large and rather sophisticated. In the interest of full disclosure, I should state that the scope of the project changed several times, at times growing and at others shrinking. Eventually, I decided for a book in three parts covering (I) foundational concepts in probability and statistics, (II) basic and commonly used advanced measurement techniques, and (III) introductory techniques in Monte Carlo simulations targeted, mostly, toward the analysis and interpretation of experimental data. As such, it became impossible to present detailed descriptions of detector technologies or the physical principles they are based on. But as it turns out, high-quality data analyses are possible even if one is not familiar with the many technical details involved in the design or construction of detectors. Detector attributes relevant for data analyses can in general be reduced to a statement of a few essential properties, and it is thus possible to carry out quality analyses without a full knowledge of all aspects of a detector's design and operation. I have thus opted to leave out detailed descriptions of detector technologies as well as particle interactions with matter and focus the discussion on some representative and illustrative examples of data calibration and analyses. Detailed discussions of detector technologies used in high-energy nuclear and particle physics may, however, be found in a plurality of graduate textbooks and technical texts. Additionally, I have also omitted few big and important topics such as interferometry (HBT), jet reconstruction, and neutral networks, for which very nice and comprehensive books or scientific reviews already exist.

Overall, this book essentially covers all basic techniques necessary for sound analyses and interpretation of experimental data. And, although it cannot cover all analysis techniques used by modern physicists, it lays a solid foundation in probability and statistics, simulation techniques, and basic measurement methods, which should equip conscientious and dedicated students with the skill set they require for a successful career in experimental nuclear or particle physics, and such that they can explore more advanced techniques on their own.

I should note, in closing, that although this book targets primarily students in nuclear and particle physicists, it should, I believe, prove to be a useful introduction to data analysis for students working in other fields, including astronomy and basically all other areas of the physical sciences. It should also, I hope, provide a useful reference for more advanced and seasoned scientists.

I would like to express my sincere acknowledgments to the many people who, through discussions and advices, have helped shape this book. These include Monika Sharma, Rosie Reed, Robert Harr, Paul Karchin, and Sergei Voloshin, who through questions and comments have helped me plan or contributed various improvements to the book. I also wish to acknowledge the important contributions of several undergraduate and graduate students, most particularly Nick Elsey, Derek Everett, Derek Hazard, Ed Kramkowski, Jinjin Pan, Jon Troyer, and Chris Zin, who served as guinea-pigs for some fractions of the material. I am grateful to my colleagues Giovanni Bonvicini, from the CLEO Collaboration, for providing a Dalitz plot; Yuri Fisyak and Zhangbu Xu, from the STAR (Solenoidal Detector at Relativistic Heavy Ion Collider [RHIC]) Collaboration, for their contribution of a dE/dx plot; and my former postdoctoral student, Sidharth Prasad, for producing exemplars of unfolding. I also acknowledge use of several sets of results from the STAR collaboration, publicly available from the collaboration's website, for the generation of figures presenting examples of flow measurements and correlation functions. I am particularly indebted to colleagues Drs. Jean Barrette, Ron Belmont, Jana Bielcikova, Panos Christakoglou, Kolja Kauder, William Llope, Prabhat Pujahari, Sidharth Prasad, Joern Putschke, and William Zajc for their detailed reading and feedback on various sections of the book corresponding to their respective areas of expertise and interest. I also wish to acknowledge Ms. Heidi Kenaga and Ms. Theresa Kornak for their meticulous proofreading of the manuscript and for being so nice in correcting my Frenglish.

Finally, I wish to acknowledge that a large fraction of the graphs and figures featured in this book were created with ROOT, Keynote, and Graphic Converter. Several of the ROOT macros I wrote for the generation of figures will be made available at the book website.

Claude A. Pruneau

How to Read This Book

Not all students, instructors, and practitioners of the field of experimental physics may have the inclination, the time, or the need to study this book in its entirety. Indeed, only a selected few may have the opportunity to read the book from cover to cover. This should not be a problem, however, because the material is organized in large blocks that are reasonably self-sufficient, and ample references to earlier or upcoming chapters, as the case may be, are included in the narrative. The book also includes a number of specialized or in-depth topics that may be skipped in a first reading. Such topics include, for instance, the formal definition of probability in §2.2, the notion of Fisher information discussed in §4.7, the technique of Kalman filtering introduced in §5.6 and for which a detailed example of application is presented in §9.2.3, as well as discussions of track and vertex reconstruction presented in §§ 9.2 and 9.3. This said, the book is designed to progressively develop and approach topics, and it should then be possible to study the material in a variety of ways, adapting the depth and breadth of coverage. The following are recommended lists of chapters and sections that should be covered given specific and targeted needs.

- Introductory course in probability and statistics:
  Chapters 2 (§§2.1, 2.3–2.11), 3 (§§3.1–3.13), 4 (§§4.1–4.6), 5 (§§5.1–5.5), 6 (§§6.1–6.6), 7 (§§7.1, 7.2, 7.4, 7.7), 13
- Advanced course in probability and statistics:
  Chapters 1, 2, 3, 4, 5, 6, 7, 13
- Introductory course in data analysis techniques (one semester):
  Chapters 1, 2 (§§2.1, 2.3–2.11), 3 (§§3.1–3.13), 4 (§§4.1–4.6), 5 (§§5.1–5.5), 6 (§§6.1–6.6), 8 (§§8.1–8.6), 9 (§§9.1, 9.2), 13
- Advanced course in data analysis techniques (two semesters):
  Chapters 1, 2 (§§2.1, 2.3–2.11), 3 (§§3.1–3.13), 4 (§§4.1–4.6), 5 (§§5.1–5.6), 6 (§§6.1–6.7), 7, 8 (§§81–8.6), 9 (§§9.1, 9.2), 12, 13, 14
- Course on correlation functions (one semester):
  Chapters 2 (§§2.5–2.13), 4 (§§4.3, 4.5, 4.6), 10, 11, 12, 13

Of course, instructors using this book should feel free to select and change the order of topics to suit their specific needs. For instance, Monte Carlo methods are formally introduced in Chapter 13 but it is often useful and inconvenient to use and discuss some of these concepts along with materials of the early chapters (e.g., Chapters 2–7).