The Probability Companion for Engineering and Computer Science

This friendly guide is the companion you need to convert pure mathematics into understanding and facility with a host of probabilistic tools. The book provides a high-level view of probability and its most powerful applications. It begins with the basic rules of probability and quickly progresses to some of the most sophisticated modern techniques in use, including Kalman filters, Monte Carlo techniques, machine learning methods, Bayesian inference and stochastic processes. It draws on 30 years of experience in applying probabilistic methods to problems in computational science and engineering, and numerous practical examples illustrate where these techniques are used in the real world. Topics of discussion range from carbon dating to Wasserstein GANs, one of the most recent developments in Deep Learning. The underlying mathematics is presented in full, but clarity takes priority over complete rigour, making this text a starting reference source for researchers and a readable overview for students.

ADAM PRÜGEL-BENNETT is Professor of Electronics and Computer Science at the University of Southampton. He received his PhD in Statistical Physics at the University of Edinburgh, where he became interested in disordered and complex systems. He currently researches in the area of mathematical modelling, optimisation and machine learning and has published many papers on these subjects.

The Probability Companion for Engineering and Computer Science

Adam Prügel-Bennett
University of Southampton

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

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/9781108480536
DOI: 10.1017/9781108635349

First published 2019

Printed in the United Kingdom by TJ International, Padstow, Cornwall

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

Library of Congress Cataloging-in-Publication Data
Names: Prügel-Bennett, Adam, 1963– author.
Title: The probability companion for engineering and computer science / Adam Prügel-Bennett, University of Southampton.
Description: Cambridge, United Kingdom ; New York, NY, USA : University Printing House, 2019. | Includes bibliographical references and index.
Identifiers: LCCN 2019015914 | ISBN 9781108480536 (hardback : alk. paper) | ISBN 9781108727709 (paperback : alk. paper)
Subjects: LCSH: Engineering–Statistical methods. | Computer science–Statistical methods. | Probabilities.
Classification: LCC TA340 .P84 2019 | DDC 519.2–dc23
LC record available at https://lccn.loc.gov/2019015914

ISBN 978-1-108-48053-6 Hardback
ISBN 978-1-108-72770-9 Paperback

Contents

Preface

Probability provides by far the most powerful and successful calculus for dealing with uncertainty. The rules of probability are reasonably quick to master; much of the interest comes from the tools and techniques that have been developed to apply probability in different areas. This book provides a high-level guide to probability theory and the tool set that has developed around it. The text started life as notes for a course aimed at research students in engineering and science. I hope the book has retained some of that original spirit. Perhaps inevitably, the book has grown and many details added. I would, however, encourage the reader not to get bogged down in the details. I believe that you can pick up the technical details when you come to use the tools to solve your problem, but it is important to have some feel for what tools are out there. The book reflects my personal interests and knowledge. No doubt there are important areas I have missed. The one benefit of my ignorance is that it keeps the book to manageable proportions. There are likely to be areas which are over-represented due to the quirks of my personal interest. I hope the balance I've struck is not too idiosyncratic and gives a reasonable overview of the practical applications of probability.

I personally dislike books that demand of their readers that they do all the problems. Consequently, I had initially intended to avoid providing exercises. In the end, however, I reconsidered when a student explained that he learns through doing. I have therefore provided exercises at the end of each chapter. Because I dislike exercises where I don't know if I have the right solution, I have supplied complete solutions to all the problems. The reader is invited to treat the exercises in any way they wish. You may want to ignore the exercises altogether, just read the solutions, or carefully work through them yourself. For those who wish to do even more exercises you may like to consult Grimmett and Stirzaker (2001b) or Mosteller (1988).

This book intentionally focuses on giving an intuitive understanding of the techniques rather than providing a mathematically rigorous treatment. I found it difficult, however, to just present formula and I have mostly tried to give complete derivations of important results. To avoid expanding the text too much I have consigned some of the technical material to appendices. I have tried to correct the text as much as I can, but I possess in abundance the human disposition to err. If errors remain (and I am sure they will), I hope they are not too off-putting. One useful lesson (though one I would prefer not to teach) is never believe things just because they are in print. This means being able to check results for consistency and derive them from first principles. Of course, it is useful to have a relatively reliable source rather than check everything from scratch. The only reward I can offer is the knowledge that I will put any corrections I receive into any new editions.

I am deeply indebted to Robert Piché from Tampere University of Technology, who as a reviewer of the book did me the great honour of providing a very detailed list of improvements and corrections. Not only did he perform the Herculean task of correcting my English, but he provided a lot of technical guidance, introducing me, for example, to a cleaner proof of Jensen's inequality, among many other significant improvements. I would also like to thank Dr Jim Bennett for carefully reading the manuscript and pointing out additional errors and confusions.

Nomenclature

$x, y, \ldots$ Normal variables are written in italics

$\boldsymbol{x}, \boldsymbol{y}, \ldots$ Vectors are written in bold roman script

$X, Y, \ldots$ Random variables are written as capitals

$\boldsymbol{X}, \boldsymbol{Y}, \ldots$ Random vectors are written as bold capitals

$\mathsf{\mathbf{M}}, \mathsf{\mathbf{A}}, \ldots$ Matrices are written as bold sans-serif capitals

$F_X(x)$ Cumulative probability function of a random variable X, page 12

$f_X(x)$ Probability density of a continuous random variable X, page 12

$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ Multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, see Equation (2.12), page 37

$\mathcal{N}(x|\mu, \sigma^2)$ Normal (or Gaussian) distribution, see Equation (2.4), page 31

$\text{Bern}(X|\mu)$ Bernoulli distribution for binary variables, see Equation (4.1), page 60

$\text{Bet}(x|a, b)$ Beta distribution, see Equation (2.7), page 34

$\text{Bin}(m|n, p)$ Binomial distribution, see Equation (2.1), page 26

$\text{Cat}(\boldsymbol{X}|\boldsymbol{p})$ Categorical distribution, see Equation (4.2), page 68

$\text{Cau}(x)$ Cauchy distribution, see Equation (2.8), page 35

$\text{Dir}(\boldsymbol{x}|\boldsymbol{\alpha})$ Dirichlet distribution, see Equation (2.13), page 38

$\text{Exp}(x|b)$ Exponential distribution, see Equation (2.6), page 32

$\text{Gam}(x|a, b)$ Gamma distribution, see Equation (2.5), page 31

$\text{Hyp}(k|N, m, n)$ Hypergeometric distribution, see Equation (2.2), page 27

$\text{LogNorm}(x|\mu, \sigma)$ Log-normal distribution, see Equation (5.2), page 85

$\text{Mult}(\boldsymbol{n}|n, \boldsymbol{p})$ Multinomial distribution, see Equation (2.10), page 36

$\text{Poi}(m|\mu)$ Poisson distribution, see Equation (2.3), page 28

$\text{U}(x|a,b)$ Uniform distribution in the interval (a,b), page 48

$\text{Wei}(x|\lambda,k)$ Weibull distribution, see Equation (2.6), page 33

$\emptyset$ The empty set

$\hat{x}$ Estimator of the quantity x, see Equation (4.1), page 61

Λ^k The $(k-1)$-dimensional unit simplex (i.e. the set of k-component vectors with non-negative elements that sum to 1), see Equation (2.9), page 36

Λ^k_n The k-dimensional discrete (integer) simplex that sums to n (i.e. the set of k non-negative integers that sum to n), see Equation (2.11), page 36

$\log(x)$ Denotes the natural logarithm of x

$\mathbb{N}$ The set of natural numbers (i.e. integers greater than 0)

$\mathbb{R}$ The set of real numbers

Ω The set of all possible elementary events

$[\![\textit{predicate}]\!]$ indicator function returning 1 if *predicate* is true and zero otherwise, see Equation (1.8), page 16

$|\mathbf{A}|$ Determinant of matrix $\mathbf{A}$, see Equation (5.2), page 97

$X \sim f_X$ The random variable X is drawn from the distribution $f_X(x)$, see Equation (3.2), page 47

$\mathbb{E}_X[g(X)]$ Expectation with respect to random variable X of some function $g(X)$, see Equation (1.6), page 15

$\mathbb{E}[g(X)]$ Short for $\mathbb{E}_X[g(X)]$ when there is no ambiguity which variable is being marginalised (averaged) over, see Equation (1.6), page 15

$\mathbb{C}\text{ov}[X,Y]$ The covariance of two random variables defined as $\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$, see Equation (1.10), page 18

$\mathbb{C}\text{ov}[\boldsymbol{X},\boldsymbol{Y}]$ The covariance matrix of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined so that the matrix $\mathbf{C} = \mathbb{C}\text{ov}[\boldsymbol{X},\boldsymbol{Y}]$ has components $C_{ij} = \mathbb{C}\text{ov}[X_i, Y_j]$, see Equation (1.10), page 18

$\mathbb{C}\text{ov}[\boldsymbol{X}]$ Short form for the covariance matrix $\mathbb{C}\text{ov}[\boldsymbol{X},\boldsymbol{X}]$, see Equation (1.10), page 18

$\mathbb{V}\text{ar}[X]$ The variance of variable X given by $\mathbb{E}[X^2] - \mathbb{E}[X]^2$, see Equation (1.8), page 17

$\neg A$ Not the event A (logical negation)

$A \vee B$ The event A or B (logical or)

$A \wedge B$ The event A and B (logical and)

$\mathbb{P}(A)$ Probability of event A happening, see Equation (1.0), page 3

$\mathbb{P}(A, B)$ Joint probability of event A and event B both happening, see Equation (1.0), page 5

$\mathbb{P}(A|B)$ Conditional probability of event A happening given event B happens, see Equation (1.1), page 6