

## 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.

Cambridge University Press  
978-1-108-48053-6 — The Probability Companion for Engineering and Computer Science  
Adam Prügel-Bennett  
Frontmatter  
[More Information](#)

---

# The Probability Companion for Engineering and Computer Science

Adam Prügel-Bennett  
*University of Southampton*



**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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108480536](http://www.cambridge.org/9781108480536)

DOI: 10.1017/9781108635349

© Adam Prügel-Bennett 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

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://lcn.loc.gov/2019015914>

ISBN 978-1-108-48053-6 Hardback

ISBN 978-1-108-72770-9 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

---

## Contents

<i>Preface</i>	xi
<i>Nomenclature</i>	xiii
<b>1 Introduction</b>	1
1.1 Why Probabilities?	2
1.2 Events and Probabilities	3
1.2.1 Events	3
1.2.2 Assigning Probabilities	5
1.2.3 Joint and Conditional Probabilities	5
1.2.4 Independence	8
1.3 Random Variables	11
1.4 Probability Densities	12
1.5 Expectations	15
1.5.1 Indicator Functions	16
1.5.2 Statistics	17
1.6 Probabilistic Inference	19
1.A Measure for Measure	23
<b>2 Survey of Distributions</b>	25
2.1 Discrete Distributions	26
2.1.1 Binomial Distribution	26
2.1.2 Hypergeometric Distribution	27
2.1.3 Poisson Distribution	28
2.2 Continuous Distributions	30
2.2.1 Normal Distribution	30
2.2.2 Gamma Distribution	31
2.2.3 Beta Distribution	34
2.2.4 Cauchy Distribution	35
2.3 Multivariate Distributions	35
2.3.1 Multinomial Distribution	35
2.3.2 Multivariate Normal Distribution	37
2.3.3 Dirichlet Distribution	37
2.4 Exponential Family	38
	v

vi		<i>Contents</i>
2.A	The Gamma Function	41
2.B	The Beta Function	43
<b>3</b>	<b>Monte Carlo</b>	<b>45</b>
3.1	Random Deviates	46
	3.1.1 Estimating Expectations	46
	3.1.2 Monte Carlo Integration	47
3.2	Uniform Random Deviates	49
3.3	Non-Uniform Random Deviates	51
	3.3.1 Transformation Method	51
	3.3.2 Rejection Sampling	54
	3.3.3 Multivariate Deviates	56
<b>4</b>	<b>Discrete Random Variables</b>	<b>59</b>
4.1	Bernoulli Trials	60
4.2	Binomial Distribution	61
	4.2.1 Statistical Properties of the Binomial Distribution	62
	4.2.2 Cumulants	64
4.3	Beyond the Binomial Distribution	66
	4.3.1 Large $n$ Limit	66
	4.3.2 Poisson Distribution	67
	4.3.3 Multinomial Distribution	68
<b>5</b>	<b>The Normal Distribution</b>	<b>74</b>
5.1	Ubiquitous Normals	75
5.2	Basic Properties	76
	5.2.1 Integrating Gaussians	76
	5.2.2 Moments and Cumulants	78
5.3	Central Limit Theorem	81
5.4	Cumulative Distribution Function of a Normal Distribution	90
5.5	Best of $n$	92
5.6	Multivariate Normal Distributions	96
5.A	Dirac Delta	103
5.B	Characteristic Function for the Cauchy Distribution	107
<b>6</b>	<b>Handling Experimental Data</b>	<b>110</b>
6.1	Estimating the Error in the Mean	111
	6.1.1 Computing Means with Errors	114
	6.1.2 Bernoulli Trials	114
6.2	Histogram	115
6.3	Significance Tests	116
6.4	Maximum Likelihood Estimate	123
6.A	Deriving the T-Distribution	126
<b>7</b>	<b>Mathematics of Random Variables</b>	<b>132</b>
7.1	Convergence	133

<i>Contents</i>	vii
7.1.1 Laws of Large Numbers	138
7.1.2 Martingales	139
7.2 Inequalities	142
7.2.1 Cauchy–Schwarz Inequality	142
7.2.2 Markov’s Inequality	143
7.2.3 Chebyshev’s Inequality	149
7.2.4 Tail Bounds and Concentration Theorems	152
7.2.5 Chernoff Bounds for Independent Bernoulli Trials	157
7.2.6 Jensen’s Inequality	159
7.2.7 Union Bound	165
7.3 Comparing Distributions	166
7.3.1 Kullback–Leibler Divergence	166
7.3.2 Wasserstein Distance	170
7.A Taylor Expansion	177
7.B Hoeffding’s Bound for Negatively Correlated Random Boolean Variables	179
7.C Constrained Optimisation	182
<b>8 Bayes</b>	187
8.1 Bayesian Statistics	188
8.2 Performing Bayesian Inference	197
8.2.1 Conjugate Priors	198
8.2.2 Uninformative Priors	205
8.2.3 Model Selection	211
8.3 Bayes with Complex Likelihoods	214
8.3.1 Hierarchical Models	214
8.3.2 MAP Solution	215
8.3.3 Variational Approximation	220
8.4 Latent Variable Models	227
8.4.1 Hidden Markov Models	235
8.4.2 Variational Auto-Encoders	240
8.5 Machine Learning	243
8.5.1 Naive Bayes Classifier	244
8.5.2 Graphical Models	245
8.A Bertrand’s Paradox	256
<b>9 Entropy</b>	258
9.1 Shannon Entropy	259
9.1.1 Information Theory	259
9.1.2 Properties of Entropy	265
9.2 Applications	267
9.2.1 Mutual Information	267
9.2.2 Maximum Entropy	267
9.2.3 The Second Law of Thermodynamics	272
9.3 Beyond Information Theory	276
9.3.1 Kolmogorov Complexity	276
9.3.2 Minimum Description Length	277
9.3.3 Fisher Information	285
9.A Functionals	290

viii		<i>Contents</i>
<b>10</b>	<b>Collective Behaviour</b>	293
10.1	Random Walk	294
10.2	Branching Processes	295
10.3	Percolation	297
10.4	Ising Model	299
10.5	Self-Organised Criticality	302
10.6	Disordered Systems	304
<b>11</b>	<b>Markov Chains</b>	308
11.1	Markov Chains	309
	11.1.1 Markov Chains and Stochastic Matrices	310
	11.1.2 Properties of Stochastic Matrices	313
	11.1.3 Ergodic Stochastic Matrices	316
11.2	Markov Chain Monte Carlo	317
	11.2.1 Detailed Balance	318
	11.2.2 Bayes and MCMC	323
	11.2.3 Convergence of MCMC	326
	11.2.4 Hybrid Monte Carlo	329
11.3	Kalman and Particle Filtering	330
	11.3.1 The Kalman Filter	332
	11.3.2 Particle Filtering	338
	11.3.3 Approximate Bayesian Computation	344
11.A	Eigenvalues and Eigenvectors of General Square Matrices	347
<b>12</b>	<b>Stochastic Processes</b>	349
12.1	Stochastic Processes	350
	12.1.1 What Are Stochastic Processes?	350
	12.1.2 Gaussian Processes	351
	12.1.3 Markov Processes	356
12.2	Diffusion Processes	359
	12.2.1 Brownian Motion	359
	12.2.2 Stochastic Differential Equations	361
	12.2.3 Fokker–Planck Equation	370
	12.2.4 Stationary Distribution of Stochastic Processes	374
	12.2.5 Pricing Options	377
12.3	Point Processes	380
	12.3.1 Poisson Processes	380
	12.3.2 Poisson Processes in One Dimension	383
	12.3.3 Chemical Reactions	385
<b>A</b>	<b>Answers to Exercises</b>	391
A.1	Answers to Chapter 1. Introduction	392
A.2	Answers to Chapter 2. Survey of Distributions	396
A.3	Answers to Chapter 3. Monte Carlo	399
A.4	Answers to Chapter 4. Discrete Random Variables	401
A.5	Answers to Chapter 5. The Normal Distribution	407
A.6	Answers to Chapter 6. Handling Experimental Data	411



<i>Contents</i>	ix
A.7 Answers to Chapter 7. Mathematics of Random Variables	414
A.8 Answers to Chapter 8. Bayes	421
A.9 Answers to Chapter 9. Entropy	427
A.10 Answers to Chapter 10. Collective Behaviour	430
A.11 Answers to Chapter 11. Markov Chains	432
A.12 Answers to Chapter 12. Stochastic Processes	441
<b>B Probability Distributions</b>	445
Table B1. Discrete Univariate Distributions	446
Table B2. Continuous Univariate Distributions	447
Table B3. Continuous and Discrete Multivariate Distributions	448
<i>Bibliography</i>	449
<i>Index</i>	454

Cambridge University Press  
978-1-108-48053-6 — The Probability Companion for Engineering and Computer Science  
Adam Prügel-Bennett  
Frontmatter  
[More Information](#)

---

## 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

- $x, y, \dots$  Normal variables are written in italics
- $\mathbf{x}, \mathbf{y}, \dots$  Vectors are written in bold roman script
- $X, Y, \dots$  Random variables are written as capitals
- $\mathbf{X}, \mathbf{Y}, \dots$  Random vectors are written as bold capitals
- $\mathbf{M}, \mathbf{A}, \dots$  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
- $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  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
- $\text{Cat}(\mathbf{X}|\mathbf{p})$  Categorical distribution, see Equation (4.2), page 68
- $\text{Cau}(x)$  Cauchy distribution, see Equation (2.8), page 35
- $\text{Dir}(\mathbf{x}|\boldsymbol{\alpha})$  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
- $\text{Mult}(\mathbf{n}|\mathbf{n}, \mathbf{p})$  Multinomial distribution, see Equation (2.10), page 36

- $\text{Poi}(m|\mu)$  Poisson distribution, see Equation (2.3), page 28
- $U(x|a, b)$  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
- $\Lambda^k$  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
- $\Lambda_n^k$  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
- $\Omega$  The set of all possible elementary events
- $\mathbb{I}[\textit{predicate}]$  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
- $\text{Cov}[X, Y]$  The covariance of two random variables defined as  $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ , see Equation (1.10), page 18
- $\text{Cov}[\mathbf{X}, \mathbf{Y}]$  The covariance matrix of two random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  defined so that the matrix  $\mathbf{C} = \text{Cov}[\mathbf{X}, \mathbf{Y}]$  has components  $C_{ij} = \text{Cov}[X_i, Y_j]$ , see Equation (1.10), page 18
- $\text{Cov}[\mathbf{X}]$  Short form for the covariance matrix  $\text{Cov}[\mathbf{X}, \mathbf{X}]$ , see Equation (1.10), page 18
- $\text{Var}[X]$  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)

*Nomenclature*

xv

$A \vee B$  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

Cambridge University Press  
978-1-108-48053-6 — The Probability Companion for Engineering and Computer Science  
Adam Prügel-Bennett  
Frontmatter  
[More Information](#)

---