Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

Bandit Algorithms

Decision-making in the face of uncertainty is a significant challenge in machine learning, and the multi-armed bandit model is a commonly used framework to address it. This comprehensive and rigorous introduction to the multi-armed bandit problem examines all the major settings, including stochastic, adversarial and Bayesian frameworks. A focus on both mathematical intuition and carefully worked proofs makes this an excellent reference for established researchers and a helpful resource for graduate students in computer science, engineering, statistics, applied mathematics and economics. Linear bandits receive special attention as one of the most useful models in applications, while other chapters are dedicated to combinatorial bandits, ranking, non-stationary problems, Thompson sampling and pure exploration. The book ends with a peek into the world beyond bandits with an introduction to partial monitoring and learning in Markov decision processes.

TOR LATTIMORE is a research scientist at DeepMind. His research is focused on decision-making in the face of uncertainty, including bandit algorithms and reinforcement learning. Before joining DeepMind, he was an assistant professor at Indiana University and a postdoctoral fellow at the University of Alberta.

CSABA SZEPESVÁRI is a professor at the Department of Computing Science of the University of Alberta and a principal investigator at the Alberta Machine Intelligence Institute. He also leads the Foundations team at DeepMind. He has co-authored a book on nonlinear approximate adaptive controllers and authored a book on reinforcement learning, in addition to publishing over 200 journal and conference papers. He is an action editor of the *Journal of Machine Learning Research*.

Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

Bandit Algorithms

Tor Lattimore
DeepMind

Csaba Szepesvári
University of Alberta

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/9781108486828

DOI: 10.1017/9781108571401

First published 2020

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: Lattimore, Tor, 1987- author. | Szepesvári, Csaba, author.
Title: Bandit algorithms / Tor Lattimore and Csaba Szepesvári.
Description: Cambridge ; New York, NY : Cambridge University Press, 2020. |
Includes bibliographical references and index.
Identifiers: LCCN 2019053276 (print) | LCCN 2019053277 (ebook) |
ISBN 9781108486828 (hardback) | ISBN 9781108571401 (epub)
Subjects: LCSH: Mathematical optimization. | Probabilities. |
Decision making–Mathematical models. | Resource allocation–Mathematical models. | Algorithms.
Classification: LCC QA402.5 .L367 2020 (print) | LCC QA402.5 (ebook) |
DDC 519.3–dc23
LC record available at https://lccn.loc.gov/2019053276
LC ebook record available at https://lccn.loc.gov/2019053277

ISBN 978-1-108-48682-8 Hardback

CAMBRIDGE
Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

Contents

Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

CAMBRIDGE
Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

Preface

Multi-armed bandits have now been studied for nearly a century. While research in the beginning was quite meandering, there is now a large community publishing hundreds of articles every year. Bandit algorithms are also finding their way into practical applications in industry, especially in on-line platforms where data is readily available and automation is the only way to scale.

We had hoped to write a comprehensive book, but the literature is now so vast that many topics have been excluded. In the end we settled on the more modest goal of equipping our readers with enough expertise to explore the specialised literature by themselves, and to adapt existing algorithms to their applications. This latter point is important. Problems in theory are all alike; every application is different. A practitioner seeking to apply a bandit algorithm needs to understand which assumptions in the theory are important and how to modify the algorithm when the assumptions change. We hope this book can provide that understanding.

What is covered in the book is covered in some depth. The focus is on the mathematical analysis of algorithms for bandit problems, but this is not a traditional mathematics book, where lemmas are followed by proofs, theorems and more lemmas. We worked hard to include guiding principles for designing algorithms and intuition for their analysis. Many algorithms are accompanied by empirical demonstrations that further aid intuition.

We expect our readers to be familiar with basic analysis and calculus and some linear algebra. The book uses the notation of measure-theoretic probability theory, but does not rely on any deep results. A dedicated chapter is included to introduce the notation and provide intuitions for the basic results we need. This chapter is unusual for an introduction to measure theory in that it emphasises the reasons to use σ-algebras beyond the standard technical justifications. We hope this will convince the reader that measure theory is an important and intuitive tool. Some chapters use techniques from information theory and convex analysis, and we devote a short chapter to each.

Most chapters are short and should be readable in an afternoon or presented in a single lecture. Some components of the book contain content that is not really about bandits. These can be skipped by knowledgeable readers, or otherwise referred to when necessary. They are marked with a (🦘) because 'Skippy the Kangaroo' skips things.[1] The same mark is used for those parts that contain useful, but perhaps overly specific information for the first-time reader. Later parts will not build on these chapters in any substantial way. Most chapters end with a list of notes and exercises. These are intended to deepen intuition and highlight

[1] Taking inspiration from Tor's grandfather-in-law, John Dillon [Anderson et al., 1977].

the connections between various subsections and the literature. There is a table of notation at the end of this preface.

Thanks

We're indebted to our many collaborators and feel privileged that there are too many of you to name. The University of Alberta, Indiana University and DeepMind have all provided outstanding work environments and supported the completion of this book. The book has benefited enormously from the proofreading efforts of a large number of our friends and colleagues. We're sorry for all the mistakes introduced after your hard work. Alphabetically, they are: Aaditya Ramdas, Abbas Mehrabian, Aditya Gopalan, Ambuj Tewari, András György, Arnoud den Boer, Branislav Kveton, Brendan Patch, Chao Tao, Christoph Dann, Claire Vernade, Emilie Kaufmann, Eugene Ji, Gellért Weisz, Gergely Neu, Johannes Kirschner, Julian Zimmert, Kwang-Sung Jun, Lalit Jain, Laurent Orseau, Marcus Hutter, Michal Valko, Omar Rivasplata, Pierre Menard, Ramana Kumar, Roman Pogodin, Ronald Ortner, Ronan Fruit, Ruihao Zhu, Shuai Li, Toshiyuki Tanaka, Wei Chen, Yoan Russac, Yufei Yi and Zhu Xiaohu. We are especially grateful to Gábor Balázs and Wouter Koolen, who both read almost the entire book. Thanks to Lauren Cowles and Cambridge University Press for providing free books for our proofreaders, tolerating the delays and for supporting a freely available PDF version. Réka Szepesvári is responsible for converting some of our school figures to their current glory. Last of all, our families have endured endless weekends of editing and multiple false promises of 'done by Christmas'. Rosina and Beáta, it really is done now!

CAMBRIDGE
Cambridge University Press
978-1-108-48682-8 — Bandit Algorithms
Tor Lattimore , Csaba Szepesvári
Frontmatter
More Information

Notation

Some sections are marked with special symbols, which are listed and described below.

This symbol is a note. Usually this is a remark that is slightly tangential to the topic at hand.

A warning to the reader.

Something important.

An experiment.

Nomenclature and Conventions

A sequence $(a_n)_{n=1}^{\infty}$ is **increasing** if $a_{n+1} \geq a_n$ for all $n \geq 1$ and **decreasing** if $a_{n+1} \leq a_n$. When the inequalities are strict, we say **strictly increasing/decreasing**. The same terminology holds for functions. We will not be dogmatic about what is the range of $\operatorname{argmin}/\operatorname{argmax}$. Sometimes they return sets, sometimes arbitrary elements of those sets and, where stated, specific elements of those sets. We will be specific when it is non-obvious/matters. The infimum of the empty set is $\inf \emptyset = \infty$ and the supremum is $\sup \emptyset = -\infty$. The empty sum is $\sum_{i \in \emptyset} a_i = 0$ and the empty product is $\prod_{i \in \emptyset} a_i = 1$.

Landau Notation

We make frequent use of the Bachmann–Landau notation. Both were nineteenth century mathematicians who could have never expected their notation to be adopted so enthusiastically by computer scientists. Given functions $f, g : \mathbb{N} \to [0, \infty)$, define

$$f(n) = O(g(n)) \Leftrightarrow \limsup_{n \to \infty} \frac{f(n)}{g(n)} < \infty,$$
$$f(n) = o(g(n)) \Leftrightarrow \lim_{n \to \infty} \frac{f(n)}{g(n)} = 0,$$

$$f(n) = \Omega(g(n)) \Leftrightarrow \liminf_{n\to\infty} \frac{f(n)}{g(n)} > 0,$$
$$f(n) = \omega(g(n)) \Leftrightarrow \liminf_{n\to\infty} \frac{f(n)}{g(n)} = \infty,$$
$$f(n) = \Theta(g(n)) \Leftrightarrow f(n) = O(g(n)) \text{ and } f(n) = \Omega(g(n)).$$

We make use of the (Bachmann–)Landau notation in two contexts. First, in proofs where limiting arguments are made, we sometimes write lower-order terms using Landau notation. For example, we might write that $f(n) = \sqrt{n} + o(\sqrt{n})$, by which we mean that $\lim_{n\to\infty} f(n)/\sqrt{n} = 1$. In this case we use the mathematical definitions as envisaged by Bachmann and Landau. The second usage is to informally describe a result without the clutter of uninteresting constants. For better or worse, this usage is often a little imprecise. For example, we will often write expressions of the form: $R_n = O(m\sqrt{dn})$. Almost always what is meant by this is that there exists a **universal constant** $c > 0$ (a constant that does not depend on either of the quantities involved) such that $R_n \le cm\sqrt{dn}$ for all (reasonable) choices of m, d and n. In this context we are careful *not* to use Landau notation to hide large lower-order terms. For example, if $f(x) = x^2 + 10^{100}x$, we will not write $f(x) = O(x^2)$, although this would be true.

Bandits

A_t	action in round t
k	number of arms/actions
n	time horizon
X_t	reward in round t
Y_t	loss in round t
π	a policy
ν	a bandit
μ_i	mean reward of arm i

Sets

$\emptyset$	empty set
$\mathbb{N}, \mathbb{N}^+$	natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$
$\mathbb{R}$	real numbers
$\bar{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$
$[n]$	$\{1, 2, 3, \dots, n-1, n\}$
2^A	the power set of set A (the set of all subsets of A)
A^*	set of finite sequences over A, $A^* = \bigcup_{i=0}^{\infty} A^i$
B_2^d	d-dimensional unit ball, $\{x \in \mathbb{R}^d : \lVert x\rVert_2 \le 1\}$
$\mathcal{P}_d$	probability simplex, $\{x \in [0,1]^{d+1} : \lVert x\rVert_1 = 1\}$
$\mathcal{P}(A)$	set of distributions over set A
$\mathfrak{B}(A)$	Borel σ-algebra on A
$[x, y]$	convex hull of vectors or real values x and y

Functions, Operators and Operations

$\lvert A\rvert$	the cardinality (number of elements) of the finite set A
$(x)^+$	$\max(x, 0)$

$a \bmod b$	remainder when natural number a is divided by b
$\lfloor x \rfloor, \lceil x \rceil$	floor and ceiling functions of x
$\mathrm{dom}(f)$	domain of function f
$\mathbb{E}$	expectation
$\mathbb{V}$	variance
Supp	support of distribution or random variable
$\nabla f(x)$	gradient of f at x
$\nabla_v f(x)$	directional derivative of f at x in direction v
$\nabla^2 f(x)$	Hessian of f at x
$\vee, \wedge$	maximum and minimum, $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$
$\mathrm{erf}(x)$	$\frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)dy$
$\mathrm{erfc}(x)$	$1 - \mathrm{erf}(x)$
$\Gamma(z)$	Gamma function, $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x)dx$
$\phi_A(x)$	support function $\phi_A(x) = \sup_{y \in A} \langle x, y \rangle$
$f^*(y)$	convex conjugate, $f^*(y) = \sup_{x \in A} \langle x, y \rangle - f(x)$
$\binom{n}{k}$	binomial coefficient
$\mathrm{argmax}_x f(x)$	maximiser or maximisers of f
$\mathrm{argmin}_x f(x)$	minimiser or minimisers of f
$\mathbb{I}\phi$	indicator function: converts Boolean ϕ into binary
$\mathbb{I}_B$	indicator of set B
$\mathrm{D}(P, Q)$	Relative entropy between probability distributions P and Q
$d(p, q)$	Relative entropy between $\mathcal{B}(p)$ and $\mathcal{B}(q)$

Linear Algebra

$e_1, \dots, e_d$	standard basis vectors of the d-dimensional Euclidean space
$\mathbf{0}, \mathbf{1}$	vectors whose elements are all zeros and all ones, respectively
$\det(A)$	determinant of matrix A
$\mathrm{trace}(A)$	trace of matrix A
$\mathrm{im}(A)$	image of matrix A
$\ker(A)$	kernel of matrix A
$\mathrm{span}(v_1, \dots, v_d)$	span of vectors $v_1, \dots, v_d$
$\lambda_{\min}(G)$	minimum eigenvalue of matrix G
$\langle x, y \rangle$	inner product, $\langle x, y \rangle = \sum_i x_i y_i$
$\\|x\\|_p$	p-norm of vector x
$\\|x\\|_G^2$	$x^\top G x$ for positive definite $G \in \mathbb{R}^{d \times d}$ and $x \in \mathbb{R}^d$
$\prec, \preceq$	Loewner partial order of positive semidefinite matrices: $A \preceq B$ ($A \prec B$) if $B - A$ is positive semidefinite (respectively, definite).

Distributions

$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{B}(p)$	Bernoulli distribution with mean p
$\mathcal{U}(a, b)$	uniform distribution supported on $[a, b]$
$\mathrm{Beta}(\alpha, \beta)$	Beta distribution with parameters $\alpha, \beta > 0$
δ_x	Dirac distribution with point mass at x

Topological

$\mathrm{cl}(A)$	closure of set A
$\mathrm{int}(A)$	interior of set A
∂A	boundary of a set A, $\partial A = \mathrm{cl}(A) \setminus \mathrm{int}(A)$
$\mathrm{co}(A)$	convex hull of A
$\mathrm{aff}(A)$	affine hull of A
$\mathrm{ri}(A)$	relative interior of A