Journal : Large 283 | Dispatch : 31-8-2018 | Pages : 2
Article No. : 9839 | ☐ LE | ☐ TYPESET
MS Code : TMIN-D-18-00165 | ☑ CP | ☑ DISK

Markov Chains and Mixing Times (Second Edition)

by David A. Levin and Yuval Peres

PROVIDENCE: AMERICAN MATHEMATICAL SOCIETY, 2017, XVI + 448 PP., US $84.00, ISBN 978-1-4704-2962-1

REVIEWED BY DAVID ALDOUS

Unlike most books reviewed in the *Mathematical Intelligencer*, this is definitely a textbook. It assumes knowledge one might acquire in the first two years of an undergraduate mathematics program—basic mathematical probability, plus linear algebra, a little graph theory, and the infamous concept of "mathematical maturity." It has the theorem–proof style of pure mathematics, but with friendly explanations of intuition and motivation.

The Topic

The topic is one major aspect of what I like to call "modern discrete probability," as opposed to balls-in-boxes-style combinatorial probability. Here a Markov chain is best envisaged as a particle jumping between a finite number of states: from state i, it jumps to state j with some specified probability p_{ij}—in other words, a random walk on an edge-weighted directed graph. A highlight of classical theory says (omitting some details here and throughout) that provided the graph is strongly connected, the time-t distribution will converge as $t \to \infty$ to a limit distribution, the *stationary distribution*. Theoretical mathematicians in the 1950s and 1960s focused on extending such general theory to countably infinite or continuous state spaces, whereas applied mathematicians at that time studied specific models such as queueing and similar operations-research-style models.

A different paradigm for finite-state chains emerged around 1980. Consider a sequence of chains on state spaces with increasing "size" parameter n, take some formalization of "close to stationarity," estimate the corresponding "mixing time" τ_n (at which the time-t distribution is close to stationarity), and study the behavior of τ_n as $n \to \infty$. In other words, study size asymptotics, not time asymptotics. A very simple example is to take binary strings of length n as states, and from a given state, we "jump" by switching a random coordinate or (with small probability) not moving. More popular examples arise from models of random card shuffling: how many shuffles are needed to make an n-card deck "well shuffled"?

But it turns out that there is a much wider range of contexts in which mixing times are relevant. The notion of a phase transition in statistical physics models such as the Ising model corresponds to whether, in the natural associated Markov chain, τ_n increases polynomially or exponentially in n. In using the Markov chain Monte Carlo (MCMC) method to reveal posterior distributions in complex big-data Bayesian statistics models, guaranteed success depends on simulating for a number of steps larger than the mixing time, and the mixing time cannot be determined by simulation but requires theory. On the mathematical side, there is an elegant and intuitively helpful connection between electrical networks and reversible (the matrix (p_{ij}) is symmetrizable) chains, described previously in the 1984 undergraduate-level monograph *Random Walks and Electric Networks*, by Doyle and Snell.

As well as applications to these preexisting topics, the new topic of approximate counting emerged in the 1980s. As a basic example, one can readily use MCMC to approximately sample uniformly from the set of colorings of a given n-vertex graph (provided the number of available colors is large enough), because one can bound the relevant mixing time. What is less obvious at first sight is that by then estimating probabilities recursively over subgraphs of size $n-1, n-2, \ldots$, one can count approximately the number of colorings of the original graph. This gives an example of a problem that is easily solved approximately via a randomized algorithm but is difficult to solve exactly by a deterministic algorithm. In the 2000s, this line of enquiry led to a program to show that computational complexity of optimization problems over random data is related to phase transitions in corresponding statistical physics models, and this remains an active area of mathematically deep research.

The Book

This is the expanded second edition (the first edition appeared in 2009) of the first and only broad-ranging yet carefully written textbook treatment of this topic. It has been used in courses in many major universities, and has ample exercises for students.

Part I of the book is "basics," illustrated by the following. One fundamental measure of "closeness to stationarity" uses variation distance, which involves the worst-case (over events) additive error in approximating the time-t distribution of a chain by the stationary probability. Upper-bounding the corresponding mixing time is most easily done, where possible, via the coupling technique. A conceptually different notion, mathematically natural in the reversible case, involves the relaxation time, defined as 1 over the spectral gap of the symmetrized matrix (p_{ij}). Techniques for upper-bounding this relaxation time involve Dirichlet forms and the distinguished path method. Also, properties of first hitting times and cover times (visiting all states) are non-obviously related to such mixing times.

Part 2 continues to more advanced issues, including the motivating contexts alluded to before. Understanding mixing times in the Ising model in detail is still an active research topic, but Chapter 15 provides basic background.

https://doi.org/10.1007/s00283-018-9839-x

Journal : Large 283 Dispatch : 31-8-2018 Pages : 2
Article No. : 9839 LE TYPESET
MS Code : TMIN-D-18-00165 CP DISK

Chains that are monotone with respect to a partial order arise quite often and have special properties, described in Chapter 22 (new). Various other special types of processes, such as lamplighter walks (Chapter 19) and the exclusion process (Chapter 23, new), are treated carefully. Also described are general theoretical developments by Peres and coauthors over the last decade, such as the "martingales and evolving sets" technique (Chapter 17) and the intriguing connections between Cesàro mixing times and hitting times on large subsets (Chapter 24, new).

The twenty-six chapters comprising over four hundred pages underline the vast range of contexts and techniques being discussed.

The Bottom Line

Taking a recently emerged topic with a massive research literature and writing a textbook that can take a student from basic undergraduate mathematics to the ability to read current research papers is a hugely impressive achievement. This book will long remain the definitive required reading for anyone wishing to engage the topic more than superficially.

Department of Statistics
367 Evans Hall
Berkeley, CA 94720
USA
e-mail: aldous@stat.berkeley.edu