Graduate Texts in Mathematics

Daniel W. Stroock

An Introduction to Markov Processes

马尔科夫过程导论

Springer

世界图书出版公司
www.wpcbj.com.cn

Daniel W. Stroock

An Introduction to Markov Processes

Springer

图书在版编目（CIP）数据

马尔科夫过程导论：英文/（美）丹尼尔斯特鲁克（Strook，D. W.）著.—北京：世界图书出版公司北京公司，2009.4

书名原文：An Introduction to Markov Processes

ISBN 978-7-5100-0448-3

Ⅰ.马… Ⅱ.丹… Ⅲ.马尔柯夫过程—研究生—教材—英文 Ⅳ.0211.62

中国版本图书馆 CIP 数据核字（2009）第 048138 号

书　　名：	An Introduction to Markov Processes
作　　者：	Daniel W. Stroock
中 译 名：	马尔科夫过程导论
责任编辑：	高蓉　刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号　100010）
联系电话：	010-64021602，010-64015659
电子信箱：	kjb@ wpcbj. com. cn
开　　本：	24 开
印　　张：	8
版　　次：	2009 年 04 月
版权登记：	图字：01-2009-0688
书　　号：	978-7-5100-0448-3/O · 663　　**定　　价：** 25.00 元

Preface

To some extent, it would be accurate to summarize the contents of this book as an intolerably protracted description of what happens when either one raises a transition probability matrix $\mathbf{P}$ (i.e., all entries $(\mathbf{P})_{ij}$ are non-negative and each row of $\mathbf{P}$ sums to 1) to higher and higher powers or one exponentiates $\mathbf{R}(\mathbf{P} - \mathbf{I})$, where $\mathbf{R}$ is a diagonal matrix with non-negative entries. Indeed, when it comes right down to it, that is all that is done in this book. However, I, and others of my ilk, would take offense at such a dismissive characterization of the theory of Markov chains and processes with values in a countable state space, and a primary goal of mine in writing this book was to convince its readers that our offense would be warranted.

The reason why I, and others of my persuasion, refuse to consider the theory here as no more than a subset of matrix theory is that to do so is to ignore the pervasive role that probability plays throughout. Namely, probability theory provides a model which both motivates and provides a context for what we are doing with these matrices. To wit, even the term "transition probability matrix" lends meaning to an otherwise rather peculiar set of hypotheses to make about a matrix. Namely, it suggests that we think of the matrix entry $(\mathbf{P})_{ij}$ as giving the probability that, in one step, a system in state i will make a transition to state j. Moreover, if we adopt this interpretation for $(\mathbf{P})_{ij}$, then we must interpret the entry $(\mathbf{P}^n)_{ij}$ of $\mathbf{P}^n$ as the probability of the same transition in n steps. Thus, as $n \to \infty$, $\mathbf{P}^n$ is encoding the long time behavior of a randomly evolving system for which $\mathbf{P}$ encodes the one-step behavior, and, as we will see, this interpretation will guide us to an understanding of $\lim_{n\to\infty}(\mathbf{P}^n)_{ij}$. In addition, and perhaps even more important, is the role that probability plays in bridging the chasm between mathematics and the rest of the world. Indeed, it is the probabilistic metaphor which allows one to formulate mathematical models of various phenomena observed in both the natural and social sciences. Without the language of probability, it is hard to imagine how one would go about connecting such phenomena to $\mathbf{P}^n$.

In spite of the propaganda at the end of the preceding paragraph, this book is written from a mathematician's perspective. Thus, for the most part, the probabilistic metaphor will be used to elucidate mathematical concepts rather than to provide mathematical explanations for non-mathematical phenomena. There are two reasons for my having chosen this perspective. First, and foremost, is my own background. Although I have occasionally tried to help people who are engaged in various sorts of applications, I have not accumulated a large store of examples which are easily translated into terms which are appropriate for a book at this level. In fact, my experience has taught me that people engaged in applications are more than competent to handle the routine problems which they encounter, and that they come to someone like me only as a last resort. As a consequence, the questions which they ask me tend to be quite difficult and the answers to those few which I can solve usually involve material which is well beyond the scope of the present book. The second reason for my writing this book in the way that I have is that I think the material itself is of sufficient interest to stand on its own. In spite of what funding agencies would have us believe, mathematics *qua* mathematics is a worthy intellectual endeavor, and I think there is a place for a modern introduction to stochastic processes which is unabashed about making mathematics its top priority.

I came to this opinion after several semesters during which I taught the introduction to stochastic processes course offered by the M.I.T. department of mathematics. The clientele for that course has been an interesting mix of undergraduate and graduate students, less than half of whom concentrate in mathematics. Nonetheless, most of the students who stay with the course have considerable talent and appreciation for mathematics, even though they lack the formal mathematical training which is requisite for a modern course in stochastic processes, at least as such courses are now taught in mathematics departments to their own graduate students. As a result, I found no ready-made choice of text for the course. On the one hand, the most obvious choice is the classic text *A First Course in Stochastic Processes*, either the original one by S. Karlin or the updated version [4] by S. Karlin and H. Taylor. Their book gives a no nonsense introduction to stochastic processes, especially Markov processes, on a countable state space, and its consistently honest, if not always easily assimilated, presentation of proofs is complemented by a daunting number of examples and exercises. On the other hand, when I began, I feared that adopting Karlin and Taylor for my course would be a mistake of the same sort as adopting Feller's book for an undergraduate introduction to probability, and this fear prevailed the first two times I taught the course. However, after using, and finding wanting, two derivatives of Karlin's classic, I took the plunge and assigned Karlin and Taylor's book. The result was very much the one which I predicted: I was far more enthusiastic about the text than were my students.

In an attempt to make Karlin and Taylor's book more palatable for the students, I started supplementing their text with notes in which I tried to couch the proofs in terms which I hoped they would find more accessible, and my efforts were rewarded with a quite positive response from my students. In fact, as my notes became more and more extensive and began to diminish the importance of the book, I decided to convert them into what is now this book, although I realize that my decision to do so may have been stupid. For one thing, the market is already close to glutted with books which purport to cover this material. Moreover, some of these books are quite popular, although my experience with them leads me to believe that their popularity is not always correlated with the quality of the mathematics they contained. Having made that pejorative comment, I will not make public which are the books which led me to this conclusion. Instead, I will only mention the books on this topic, besides Karlin and Taylor's, which I very much liked. Namely,

J. Norris's book [5] is an excellent introduction to Markov processes which, at the same time, provides its readers with a good place to exercise their measure-theoretic skills. Of course, Norris's book is only appropriate for students who have measure-theoretic skills to exercise. On the other hand, for students who possess those skills, Norris's book is a place where they can see measure theory put to work in an attractive way. In addition, Norris has included many interesting examples and exercises which illustrate how the subject can be applied. The present book includes most of the mathematical material contained in [5], but the proofs here demand much less measure theory than his do. In fact, although I have systematically employed measure theoretic terminology (Lebesgue's Dominated Convergence Theorem, the Monotone Convergence Theorem, etc.), which is explained in Chapter 6, I have done so only to familiarize my readers with the jargon which they will encounter if they delve more deeply into the subject. In fact, because the state spaces in this book are countable, the applications which I have made of Lebesgue's theory are, with one notable exception, entirely trivial. The one exception, which is made in § 6.2, is that I have included a proof that there exist countably infinite families of mutually independent random variables. Be that as it may, the reader who is ready to accept that such families exist has no need to consult Chapter 6 except for terminology and the derivation of a few essentially obvious facts about series. For more advanced students, an excellent treatment of Markov chains on a general state space can be found in the book [6] by D. Revuz.

The organization of this book should be more or less self-evident from the table of contents. In Chapter 1, I give a bare hands treatment of the basic facts, with particular emphasis on recurrence and transience, about nearest neighbor random walks on the square, d-dimensional lattice $\mathbb{Z}^d$. Chapter 2 introduces the study of ergodic properties, and this becomes the central theme which ties together Chapters 2 through 5. In Chapter 2, the systems under consideration are Markov chains (i.e., the time parameter is discrete), and the driving force behind the development there is an idea which was introduced by Doeblin. Restricted as the applicability of Doeblin's idea may be, it has the enormous advantage over the material in Chapters 3 and 4 that it provides an estimate on the rate at which the chain is converging to its equilibrium distribution. After giving a reasonably thorough account of Doeblin's theory, in Chapter 3 I study the ergodic properties of Markov chains which do not necessarily satisfy Doeblin's condition. The main result here is the one summarized in equation (3.2.15). Even though it is completely elementary, the derivation of (3.2.15), is, without doubt, the most demanding piece of analysis in the entire book. So far as I know, every proof of (3.2.15) requires work at some stage. In supposedly "simpler" proofs, the work is hidden elsewhere (either measure theory, as in [5] and [6], or in operator theory, as in [2]). The treatment given here, which is a re-working of the one in [4] based on Feller's renewal theorem, demands nothing more of the reader than a thorough understanding of arguments involving limits superior, limits inferior, and their role in proving that limits exist. In Chapter 4, Markov chains are replaced by continuous-time Markov processes (still on a countable state space). I do this first in the case when the rates are bounded and therefore problems of possible explosion do not arise. Afterwards, I allow for unbounded rates and develop criteria, besides boundedness, which guarantee non-explosion. The remainder of the chapter is devoted to transferring the results obtained for Markov chains in Chapter 3 to the continuous-time setting. Aside from Chapter 6, which is more like an appendix than an integral part of the book, the book ends with Chapter 5. The goal in Chapter 5 is to obtain quantitative results, reminiscent of, if not as strong as, those in Chapter 2, when Doeblin's theory either fails entirely or yields rather poor estimates. The new ingredient in Chapter 5 is the assumption that the chain or process is reversible (i.e., the transition probability is self-adjoint in the L^2-space of its stationary distribution), and the engine which makes everything go is the associated Dirichlet form. In the final section, the power of the Dirichlet form methodology is tested in an analysis of the Metropolis (a.k.a. as simulated annealing) algorithm. Finally, as I said before, Chapter 6 is an appendix in which the ideas and terminology of Lebesgue's theory of measure and integration are reviewed. The one substantive part of Chapter 6 is the construction, alluded to earlier, in § 6.2.1.

Finally, I have reached the traditional place reserved for thanking those individuals who, either directly or indirectly, contributed to this book. The principal direct contributors are the many students who suffered with various and spontaneously changing versions of this book. I am particularly grateful to Adela Popescu whose careful reading brought to light many minor and a few major errors which have been removed and, perhaps, replaced by new ones. Thanking, or even identifying, the indirect contributors is trickier. Indeed, they include all the individuals, both dead and alive, from whom I received my education, and I am not about to bore you with even a partial list of who they were or are. Nonetheless, there is one person who, over a period of more than ten years, patiently taught me to appreciate the sort of material treated here. Namely, Richard A. Holley, to whom I have dedicated this book, is a *true probabilist.* To wit, for Dick, intuitive understanding usually precedes his mathematically rigorous comprehension of a probabilistic phenomenon. This statement should lead no one to to doubt Dick's powers as a rigorous mathematician. On the contrary, his intuitive grasp of probability theory not only enhances his own formidable mathematical powers, it has saved me and others from blindly pursuing flawed lines of reasoning. As all who have worked with him know, reconsider what you are saying if ever, during some diatribe into which you have launched, Dick quietly says "I don't follow that."

In addition to his mathematical prowess, every one of Dick's many students will attest to his wonderful generosity. I was not his student, but I was his colleague, and I can assure you that his generosity is not limited to his students.

Daniel W. Stroock, August 2004

Daniel W. Stroock
MIT
Department of Mathematics, Rm. 272
Massachusetts Ave 77
02139-4307 Cambridge, USA
dws@math.mit.edu

Mathematics Subject Classification (2000): 60-01, 60J10, 60J27

ISSN 0072-5285
ISBN 3-540-23499-3 Springer Berlin Heidelberg New York

Library of Congress Control Number: 2004113930

Springer is a part of Springer Science+Business Media

springeronline.com

Contents

CHAPTER 1

Random Walks
A Good Place to Begin

The purpose of this chapter is to discuss some examples of Markov processes which can be understood even before the term "Markov process" is. Indeed, anyone who has been introduced to probability theory will recognize that these processes all derive from consideration of elementary "coin tossing."

1.1 Nearest Neighbor Random Walks on $\mathbb{Z}$

Let p be a fixed number from the open interval $(0,1)$, and suppose that[1] $\{B_n : n \in \mathbb{Z}^+\}$ is a sequence of $\{-1,1\}$-valued, identically distributed *Bernoulli random variables*[2] which are 1 with probability p. That is, for any $n \in \mathbb{Z}^+$ and any $E \equiv (\epsilon_1, \dots, \epsilon_n) \in \{-1,1\}^n$,

$$\begin{aligned} &\mathbb{P}\big(B_1 = \epsilon_1, \dots, B_n = \epsilon_n\big) = p^{N(E)} q^{n-N(E)} \text{ where } q \equiv 1-p \text{ and} \\ &N(E) \equiv \#\{m : \epsilon_m = 1\} = \frac{n + S_n(E)}{2} \quad \text{when } S_n(E) \equiv \sum_1^n \epsilon_m. \end{aligned} \tag{1.1.1}$$

Next, set

$$X_0 = 0 \quad \text{and} \quad X_n = \sum_{m=1}^n B_m \quad \text{for } n \in \mathbb{Z}^+. \tag{1.1.2}$$

The existence of the family $\{B_n : n \in \mathbb{Z}^+\}$ is the content of § 6.2.1.

The above family of random variables $\{X_n : n \in \mathbb{N}\}$ is often called a *nearest neighbor random walk* on $\mathbb{Z}$. Nearest neighbor random walks are examples of Markov processes, but the description which we have just given is the one which would be given in elementary probability theory, as opposed to a course, like this one, devoted to stochastic processes. Namely, in the study of stochastic processes the description should emphasize the dynamic aspects

[1] $\mathbb{Z}$ is used to denote the set of all integers, of which $\mathbb{N}$ and $\mathbb{Z}^+$ are, respectively, the non-negative and positive members.

[2] For historical reasons, mutually independent random variables which take only two values are often said to be Bernoulli random variables.

of the family. Thus, a stochastic process oriented description might replace (1.1.2) by

$$\mathbb{P}(X_0 = 0) = 1 \text{ and}$$

(1.1.3)
$$\mathbb{P}\big(X_n - X_{n-1} = \epsilon \,\big|\, X_0, \dots, X_{n-1}\big) = \begin{cases} p & \text{if } \epsilon = 1 \\ q & \text{if } \epsilon = -1, \end{cases}$$

where $\mathbb{P}\big(X_n - X_{n-1} = \epsilon \,\big|\, X_0, \dots, X_{n-1}\big)$ denotes the *conditional probability* (cf. § 6.4.1) that $X_n - X_{n-1} = \epsilon$ given $\sigma(\{X_0, \dots, X_{n-1}\})$. Notice that (1.1.3) is indeed more dynamic a description than the one in (1.1.2). Specifically, it says that the process starts from 0 at time $n = 0$ and proceeds so that, at each time $n \in \mathbb{Z}^+$, it moves one step forward with probability p or one step backward with probability q, independent of where it has been before time n.

1.1.1. Distribution at Time n: In this subsection, we will present two approaches to computing $\mathbb{P}(X_n = m)$. The first computation is based on the description given in (1.1.2). Namely, from (1.1.2) it is clear that $\mathbb{P}(|X_n| \le n) = 1$. In addition, it is clear that

$$n \text{ odd} \implies \mathbb{P}(X_n \text{ is odd}) = 1 \quad \text{and} \quad n \text{ even} \implies \mathbb{P}(X_n \text{ is even}) = 1.$$

Finally, given $m \in \{-n, \dots, n\}$ with the same parity as n and a string $E = (\epsilon_1, \dots, \epsilon_n) \in \{-1, 1\}^n$ with (cf. (1.1.1)) $S_n(E) = m$, $N(E) = \frac{n+m}{2}$ and so

$$\mathbb{P}\big(B_1 = \epsilon_1, \dots, B_n = \epsilon_n\big) = p^{\frac{n+m}{2}} q^{\frac{n-m}{2}}.$$

Hence, because, when $\binom{\ell}{k} \equiv \frac{\ell!}{k!(\ell-k)!}$ is the *binomial coefficient* "ℓ choose k," there are $\binom{n}{\frac{m+n}{2}}$ such strings E, we see that

(1.1.4)
$$\mathbb{P}(X_n = m) = \binom{n}{\frac{m+n}{2}} p^{\frac{n+m}{2}} q^{\frac{n-m}{2}}$$
$$\text{if } m \in \mathbb{Z},\ |m| \le n, \text{ and } m \text{ has the same parity as } n$$

and is 0 otherwise.

Our second computation of the same probability will be based on the more dynamic description given in (1.1.3). To do this, we introduce the notation $(P^n)_m \equiv \mathbb{P}(X_n = m)$. Obviously, $(P^0)_m = \delta_{0,m}$, where $\delta_{k,\ell}$ is the *Kronecker symbol* which is 1 when $k = \ell$ and 0 otherwise. Further, from (1.1.3), we see that $\mathbb{P}(X_n = m)$ equals

$$\begin{aligned}&\mathbb{P}\big(X_{n-1} = m-1 \;\&\; X_n = m\big) + \mathbb{P}\big(X_{n-1} = m+1 \;\&\; X_n = m\big)\\ &\quad = p\mathbb{P}\big(X_{n-1} = m-1\big) + q\mathbb{P}\big(X_{n-1} = m+1\big).\end{aligned}$$

That is,

(1.1.5)
$$(P^0)_m = \delta_{0,m} \quad \text{and} \quad (P^n)_m = p(P^{n-1})_{m-1} + q(P^{n-1})_{m+1}.$$

Obviously, (1.1.5) provides a complete, albeit implicit, prescription for computing the numbers $(P^n)_m$, and one can easily check that the numbers given by (1.1.4) satisfy this prescription. Alternatively, one can use (1.1.5) plus induction on n to see that $(P^n)_m = 0$ unless $m = 2\ell - n$ for some $0 \le \ell \le n$ and that $(C^n)_\ell = (C^n)_{\ell-1} + (C^n)_{\ell-1}$ when $(C^n)_\ell \equiv p^{-\ell}q^{n-\ell}(P^n)_{2\ell-n}$. In other words, the coefficients $\{(C^n)_\ell : n \in \mathbb{N} \ \& \ 0 \le \ell \le n\}$ are given by Pascal's triangle and are therefore the binomial coefficients.

1.1.2. Passage Times via the Reflection Principle: More challenging than the computation in §1.1.1 is finding the distribution of the first passage time to a point $a \in \mathbb{Z}$. That is, given $a \in \mathbb{Z} \setminus \{0\}$, set[3]

$$(1.1.6) \qquad \zeta_a = \inf\{n \ge 1 : X_n = a\} \ (\equiv \infty \text{ when } X_n \ne a \text{ for any } n \ge 1).$$

Then ζ_a is the *first passage time* to a, and our goal here is to find its distribution. Equivalently, we want an expression for $\mathbb{P}(\zeta_a = n)$, and clearly, by the considerations in §1.1.1, we need only worry about n's which satisfy $n \ge |a|$ and have the same parity as a.

Again we will present two approaches to this problem, here based on (1.1.2) and in §1.1.5 on (1.1.3). To carry out the one based on (1.1.2), assume that $a \in \mathbb{Z}^+$, suppose that $n \in \mathbb{Z}^+$ has the same parity as a, and observe first that

$$\mathbb{P}(\zeta_a = n) = \mathbb{P}\big(X_n = a \ \& \ \zeta_a > n-1\big) = p\mathbb{P}\big(\zeta_a > n-1 \ \& \ X_{n-1} = a-1\big).$$

Hence, it suffices for us to compute $\mathbb{P}\big(\zeta_a > n-1 \ \& \ X_{n-1} = a-1\big)$. For this purpose, note that for any $E \in \{-1,1\}^{n-1}$ with $S_{n-1}(E) = a-1$, the event $\{(B_1, \dots, B_{n-1}) = E\}$ has probability $p^{\frac{n+a}{2}-1}q^{\frac{n-a}{2}}$. Thus,

$$(*) \qquad \mathbb{P}(\zeta_a = n) = \mathcal{N}(n,a)p^{\frac{n+a}{2}}q^{\frac{n-a}{2}}$$

where $\mathcal{N}(n,a)$ is the number of $E \in \{-1,1\}^{n-1}$ with the properties that $S_\ell(E) \le a-1$ for $0 \le \ell \le n-1$ and $S_{n-1}(E) = a-1$. That is, everything comes down to the computation of $\mathcal{N}(n,a)$. Alternatively, since $\mathcal{N}(n,a) = \binom{n-1}{\frac{n+a}{2}-1} - \mathcal{N}'(n,a)$, where $\mathcal{N}'(n,a)$ is the number of $E \in \{-1,1\}^{n-1}$ such that $S_{n-1}(E) = a-1$ and $S_\ell(E) \ge a$ for some $\ell \le n-1$, we need only compute $\mathcal{N}'(n,a)$. For this purpose we will use a beautiful argument known as the *reflection principle*. Namely, consider the set $P(n,a)$ of paths $(S_0, \dots, S_{n-1}) \in \mathbb{Z}^n$ with the properties that $S_0 = 0$, $S_\ell - S_{m-1} \in \{-1,1\}$ for $1 \le m \le n-1$, and $S_m \ge a$ for some $1 \le m \le n-1$. Clearly, $\mathcal{N}'(n,a)$ is the numbers of paths in the set $L(n,a)$ consisting of those $(S_0, \dots, S_{n-1}) \in P(n,a)$ for which $S_{n-1} = a-1$, and, as an application of the reflection principle, we will show that the set $L(n,a)$ has the same number of elements as the set $U(n,a)$ whose elements are those paths $(S_0, \dots, S_{n-1}) \in P(n,a)$ for which $S_{n-1} = a+1$. Since $(S_0, \dots, S_{n-1}) \in U(n,a)$ if and only if $S_0 = 0$, $S_m - S_{m-1} \in \{-1,1\}$

[3] As the following indicates, we take the infemum over the empty set to be $+\infty$.

for all $1 \le m \le n-1$, and $S_{n-1} = a+1$, we already know how to count them: there are $\binom{n-1}{\frac{n+a}{2}}$ of them. Hence, all that remains is to provide the advertised application of the reflection principle. To this end, for a given $\mathbf{S} = (S_0, \dots, S_{n-1}) \in P(n,a)$, let $\ell(\mathbf{S})$ be the smallest $0 \le k \le n-1$ for which $S_k \ge a$, and define the *reflection* $\mathfrak{R}(\mathbf{S}) = (\hat{S}_0, \dots, \hat{S}_{n-1})$ of $\mathbf{S}$ so that $\hat{S}_m = S_m$ if $0 \le m \le \ell(\mathbf{S})$ and $\hat{S}_k = 2a - S_k$ if $\ell(\mathbf{S}) < m \le n-1$. Clearly, $\mathfrak{R}$ maps $L(n,a)$ into $U(n,a)$ and $U(n,a)$ into $L(n,a)$. In addition, $\mathfrak{R}$ is idempotent: its composition with itself is the identity map. Hence, as a map from $L(n,a)$ to $U(n,a)$, $\mathfrak{R}$ it must be both one-to-one and onto, and so $L(n,a)$ and $U(n,a)$ have the same numbers of elements.

We have now shown that $\mathcal{N}'(n,a) = \binom{n-1}{\frac{n+a}{2}}$ and therefore that

$$\mathcal{N}(n,a) = \binom{n-1}{\frac{n+a}{2}-1} - \binom{n-1}{\frac{n+a}{2}}.$$

Finally, after plugging this into (*), we arrive at

$$\mathbb{P}(\zeta_a = n) = \left[\binom{n-1}{\frac{n+a}{2}-1} - \binom{n-1}{\frac{n+a}{2}}\right] p^{\frac{n+a}{2}} q^{\frac{n-a}{2}},$$

which simplifies to the remarkably simple expression

$$\mathbb{P}(\zeta_a = n) = \frac{a}{n}\binom{n}{\frac{n+a}{2}} p^{\frac{n+a}{2}} q^{\frac{n-a}{2}} = \frac{a}{n}\mathbb{P}(X_n = a).$$

The computation when $a < 0$ can be carried out either by repeating the argument just given or, after reversing the roles of p and q, applying the preceding result to $-a$. However one arrives at it, the general result is that

$$\text{(1.1.7)} \qquad a \ne 0 \implies \mathbb{P}(\zeta_a = n) = \frac{|a|}{n}\binom{n}{\frac{n+a}{2}} p^{\frac{n+a}{2}} q^{\frac{n-a}{2}} = \frac{|a|}{n}\mathbb{P}(X_n = a)$$

for $n \ge |a|$ with the same parity as a and is 0 otherwise.

1.1.3. Some Related Computations: Although the formula in (1.1.7) is elegant, it is not particularly transparent. In particular, it is not at all evident how one can use it to determine whether $\mathbb{P}(\zeta_a < \infty) = 1$. To carry out this computation, let $a > 0$ be given, and write of $\zeta_a = f_a(B_1, \dots, B_n, \dots)$, where f_a is the function which maps $\{-1,1\}^{\mathbb{Z}^+}$ into $\mathbb{Z}^+ \cup \{\infty\}$ so that, for each $n \in \mathbb{N}$,

$$f_a(\epsilon_1, \dots, \epsilon_n, \dots) > n \iff \sum_{\ell=1}^{m} \epsilon_\ell < a \quad \text{for } 1 \le m \le n.$$

Because the event $\{\zeta_a = m\}$ depends only on $(B_1, \dots, B_m)$ and

$$\text{(1.1.8)} \qquad \begin{gathered} \zeta_a = m \implies \zeta_{a+1} = m + \zeta_1 \circ \Sigma^m \\ \text{where } \zeta_1 \circ \Sigma^m \equiv f_1(B_{m+1}, \dots, B_{m+n}, \dots), \end{gathered}$$

$\{\zeta_a = m \ \& \ \zeta_{a+1} < \infty\} = \{\zeta_a = m\} \cap \{\zeta_1 \circ \Sigma^m < \infty\}$, and $\{\zeta_a = m\}$ is independent of $\{\zeta_1 \circ \Sigma^m < \infty\}$. In particular, this leads to

$$\begin{aligned}\mathbb{P}(\zeta_{a+1} < \infty) &= \sum_{m=1}^{\infty} \mathbb{P}(\zeta_a = m \ \& \ \zeta_{a+1} < \infty) \\ &= \sum_{m=1}^{\infty} \mathbb{P}(\zeta_a = m)\mathbb{P}(\zeta_1 \circ \Sigma^m < \infty) \\ &= \mathbb{P}(\zeta_1 < \infty) \sum_{m=1}^{\infty} \mathbb{P}(\zeta_a = m) = \mathbb{P}(\zeta_1 < \infty)\mathbb{P}(\zeta_a < \infty),\end{aligned}$$

since $(B_{m+1}, \dots, B_{m+n}, \dots)$ and $(B_1, \dots, B_n, \dots)$ have the same distribution and therefore so do $\zeta_1 \circ \Sigma^m$ and ζ_1. The same reasoning applies equally well when $a < 0$, only now with -1 playing the role of 1. In other words, we have proved that

$$\text{(1.1.9)} \qquad \mathbb{P}(\zeta_a < \infty) = \mathbb{P}\big(\zeta_{\mathrm{sgn}(a)} < \infty\big)^{|a|} \quad \text{for } a \in \mathbb{Z} \setminus \{0\},$$

where $\mathrm{sgn}(a)$, the *signum* of a, is 1 or -1 according to whether $a > 0$ or $a < 0$. In particular, this shows that $\mathbb{P}(\zeta_1 < \infty) = 1 \implies \mathbb{P}(\zeta_a < \infty) = 1$ and $\mathbb{P}(\zeta_{-1} < \infty) = 1 \implies \mathbb{P}(\zeta_{-a} < \infty) = 1$ for all $a \in \mathbb{Z}^+$.

In view of the preceding, we need only look at $\mathbb{P}(\zeta_1 < \infty)$. Moreover, by the Monotone Convergence Theorem, Theorem 6.1.9,

$$\mathbb{P}(\zeta_1 < \infty) = \lim_{s \nearrow 1} \mathbb{E}\big[s^{\zeta_1}\big] = \lim_{s \nearrow 1} \sum_{n=1}^{\infty} s^{2n-1}\mathbb{P}(\zeta_1 = 2n-1).$$

Applying (1.1.7) with $a = 1$, we know that

$$\mathbb{P}(\zeta_1 = 2n-1) = \frac{1}{2n-1}\binom{2n-1}{n} p^n q^{n-1}.$$

Next, note that

$$\begin{aligned}\frac{1}{2n-1}\binom{2n-1}{n} &= \frac{(2(n-1))!}{n!(n-1)!} = \frac{2^{n-1}}{n!}\prod_{m=1}^{n-1}(2m-1) \\ &= \frac{4^{n-1}}{n!}\prod_{m=1}^{n-1}\left(m - \tfrac{1}{2}\right) = (-1)^{n-1}\frac{4^n}{2}\binom{\frac{1}{2}}{n},\end{aligned}$$

where[4], for any $\alpha \in \mathbb{R}$,

$$\binom{\alpha}{n} \equiv \begin{cases} 1 & \text{if } n = 0 \\ \frac{1}{n!}\prod_{m=0}^{n-1}(\alpha - m) & \text{if } n \in \mathbb{Z}^+ \end{cases}$$

[4] In the preceding, we have adopted the convention that $\prod_{j=k}^{\ell} a_j = 1$ if $\ell < k$.

is the *generalized binomial coefficient* which gives the coefficient of x^n in the Taylor's expansion of $(1+x)^\alpha$ around $x = 0$. Hence,

$$\sum_{n=1}^{\infty} s^{2n-1}\mathbb{P}(\zeta_1 = 2n-1) = -\frac{1}{2qs}\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}(-4pqs^2)^n = \frac{1-\sqrt{1-4pqs^2}}{2qs},$$

and so

$$\mathbb{E}\big[s^{\zeta_1}\big] = \frac{1-\sqrt{1-4pqs^2}}{2qs} \quad \text{for } |s|<1. \tag{1.1.10}$$

Of course, by symmetry, one can reverse the roles of p and q to obtain

$$\mathbb{E}\big[s^{\zeta_{-1}}\big] = \frac{1-\sqrt{1-4pqs^2}}{2ps} \quad \text{for } |s|<1. \tag{1.1.11}$$

By letting $s \nearrow 1$ in (1.1.10) and noting that $1-4pq = (p+q)^2 - 4pq = (p-q)^2$, we see that[5]

$$\lim_{s\nearrow 1}\mathbb{E}\big[s^{\zeta_1}\big] = \frac{1-|p-q|}{2q} = \frac{p\wedge q}{q},$$

and so

$$\mathbb{P}(\zeta_1 < \infty) = \begin{cases} 1 & \text{if } p \ge q \\ \frac{p}{q} & \text{if } p < q. \end{cases}$$

Of course, $\mathbb{P}(\zeta_{-1} < \infty)$ is given by the same formula, only with the roles of p and q reversed. Thus,

$$\mathbb{P}(\zeta_a < \infty) = \begin{cases} 1 & \text{if } a \in \mathbb{Z}^+ \ \&\ p \ge q \text{ or } -a \in \mathbb{Z}^+ \ \&\ p \le q \\ \left(\frac{p}{q}\right)^a & \text{if } a \in \mathbb{Z}^+ \ \&\ p < q \text{ or } -a \in \mathbb{Z}^+ \ \&\ p > q. \end{cases} \tag{1.1.12}$$

1.1.4. Time of First Return: Having gone to so much trouble to arrive at (1.1.12), it is only reasonable to draw from it a famous conclusion about the *recurrence* properties of nearest neighbor random walks on $\mathbb{Z}$. Namely, let

$$\rho_0 \equiv \inf\{n \ge 1 : X_n = 0\} \ (\equiv \infty \text{ if } X_n \ne 0 \text{ for all } n \ge 1)$$

be the *time of first return* to 0. Then, by precisely the same sort of reasoning which allowed us to arrive at (1.1.9), we see that $\mathbb{P}(X_1 = 1 \ \&\ \rho_0 < \infty) = p\mathbb{P}(\zeta_{-1} < \infty)$ and $\mathbb{P}(X_1 = -1 \ \&\ \rho_0 < \infty) = q\mathbb{P}(\zeta_1 < \infty)$, and so, by (1.1.12),

$$\mathbb{P}(\rho_0 < \infty) = 2(p \wedge q). \tag{1.1.13}$$

[5] We use $a \wedge b$ to denote the minimum $\min\{a,b\}$ of $a, b \in \mathbb{R}$.