Lecture Notes in Mathematics

Amir Dembo
Tadahisa Funaki

Lectures on Probability Theory and Statistics

1869

Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003

Editor: J. Picard

Springer

A. Dembo · T. Funaki

Lectures on Probability Theory and Statistics

Ecole d'Eté de Probabilités
de Saint-Flour XXXIII - 2003

Editor: Jean Picard

Springer

Authors and Editors

Amir Dembo
Mathematics Department
Stanford University
Stanford, CA 94305
USA
e-mail: amir@math.stanford.edu

Tadahisa Funaki
Graduate School of Mathematical Sciences
University of Tokyo
Komaba
Tokyo 153-8914
Japan
e-mail: funaki@ms.u-tokyo.ac.jp

Jean Picard
Laboratoire de Mathématiques Appliquées
UMR CNRS 6620
Université Blaise Pascal (Clermont-Ferrand)
63177 Aubière Cedex
France
e-mail: jean.picard@math.univ-bpclermont.fr

Cover: Blaise Pascal (1623-1662)

Library of Congress Control Number: 2005931754

Mathematics Subject Classification (2000): 60J10, 60J65, 60K35, 28A80, 28A78, 31C15, 82C41
82B24, 82B31, 82B41, 82C24, 82C31, 82C41, 35J20, 35K55, 35R35

ISSN print edition: 0075-8434
ISSN electronic edition: 1617-9692
ISSN Ecole d'Eté de Probabilités de St. Flour, print edition: 0721-5363
ISBN-10 3-540-26069-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-26069-1 Springer Berlin Heidelberg New York
DOI 10.1007/b136622

Typesetting: by the authors and TechBooks using a Springer LaTeX package

Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 11429579 41/TechBooks 5 4 3 2 1 0

Lecture Notes in Mathematics

Edited by J.-M. Morel, F. Takens and B. Teissier

Editorial Policy for Multi-Author Publications: Summer Schools / Intensive Courses

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications - quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome. Manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They should provide sufficient motivation, examples and applications. There should also be an introduction making the text comprehensible to a wider audience. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this „lecture notes" character.

2. In general SUMMER SCHOOLS and other similar INTENSIVE COURSES are held to present mathematical topics that are close to the frontiers of recent research to an audience at the beginning or intermediate graduate level, who may want to continue with this area of work, for a thesis or later. This makes demands on the didactic aspects of the presentation. Because the subjects of such schools are advanced, there often exists no textbook, and so ideally, the publication resulting from such a school could be a first approximation to such a textbook.

   Usually several authors are involved in the writing, so it is not always simple to obtain a unified approach to the presentation.

   For prospective publication in LNM, the resulting manuscript should not be just a collection of course notes, each of which has been developed by an individual author with little or no co-ordination with the others, and with little or no common concept. The subject matter should dictate the structure of the book, and the authorship of each part or chapter should take secondary importance. Of course the choice of authors is crucial to the quality of the material at the school and in the book, and the intention here is not to belittle their impact, but simply to say that the book should be planned to be written by these authors jointly, and not just assembled as a result of what these authors happen to submit.

   This represents considerable preparatory work (as it is imperative to ensure that the authors know these criteria before they invest work on a manuscript), and also considerable editing work afterwards, to get the book into final shape. Still it is the form that holds the most promise of a successful book that will be used by its intended audience, rather than yet another volume of proceedings for the library shelf.

3. Manuscripts should be submitted (preferably in duplicate) either to Springer's mathematics editorial in Heidelberg, or to one of the series editors (with a copy to Springer). Volume editors are expected to arrange for the refereeing, to the usual scientific standards, of the individual contributions. If the resulting reports can be forwarded to us (series editors or Springer) this is very helpful. If no reports are forwarded or if other questions remain unclear in respect of homogeneity etc, the series editors may wish to consult external referees for an overall evaluation of the volume. A final decision to publish can be made only on the basis of the complete manuscript; however a preliminary decision can be based on a pre-final or incomplete manuscript. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter.

   Volume editors and authors should be aware that incomplete or insufficiently close to final manuscripts almost always result in longer evaluation times. They should also be aware that parallel submission of their manuscript to another publisher while under consideration for LNM will in general lead to immediate rejection.

Continued on inside back-cover

Lecture Notes in Mathematics 1869

Subseries:
Ecole d'Eté de Probabilités de Saint-Flour

Preface

Three series of lectures were given at the 33rd Probability Summer School in Saint-Flour (July 6–23, 2003), by the Professors Dembo, Funaki and Massart. This volume contains the courses of Professors Dembo and Funaki. The course of Professor Massart, entitled "Concentration inequalities and model selection", will appear in another volume. We are grateful to the authors for their important contribution.

64 participants have attended this school. 31 of them have given a short lecture. The lists of participants and of short lectures are enclosed at the end of the volume.

The Saint-Flour Probability Summer School was founded in 1971. Here are the references of Springer volumes where lectures of previous years were published. All numbers refer to the *Lecture Notes in Mathematics* series, except S-50 which refers to volume 50 of the *Lecture Notes in Statistics* series.

1971: vol 307	1980: vol 929	1990: vol 1527	1997: vol 1717
1973: vol 390	1981: vol 976	1991: vol 1541	1998: vol 1738
1974: vol 480	1982: vol 1097	1992: vol 1581	1999: vol 1781
1975: vol 539	1983: vol 1117	1993: vol 1608	2000: vol 1816
1976: vol 598	1984: vol 1180	1994: vol 1648	2001: vol 1837 & 1851
1977: vol 678	1985/86/87: vol 1362 & S-50		2002: vol 1840
1978: vol 774	1988: vol 1427	1995: vol 1690	
1979: vol 876	1989: vol 1464	1996: vol 1665	

Further details can be found on the summer school web site
`http://math.univ-bpclermont.fr/stflour/`

Jean Picard
Clermont-Ferrand, September 2005

Contents

Part I Amir Dembo: Favorite Points, Cover Times and Fractals

Part II Tadahisa Funaki: Stochastic Interface Models

Part I

Amir Dembo: Favorite Points, Cover Times and Fractals

Favorite Points, Cover Times and Fractals

Amir Dembo

Department of Mathematics, Stanford University, Stanford, CA 94305,
amir@math.stanford.edu

1 Overview

In this course we follow recent advances in the study of the fractal nature of certain random sets, emphasizing the methods used to obtain such results. We focus on some of the fine properties of the sample path of the most basic stochastic processes such as simple random walks, Brownian motion, and symmetric stable processes. As we shall see, probability on trees inspires many of our proofs, with trees used to model the relevant correlation structure. Along the way we also mention quite a few challenging open research problems. Among the methods that will be detailed here are

- Cover time for Markov chains (see Chap. 2).
- The dimension of discrete lim sup random fractals (see Chap. 3).
- The truncated second moment method (see Chap. 4).
- The KMT strong approximation construction (see Chap. 5).
- Ciesielski-Taylor identities (see Chap. 6).

The highly recommended survey of Taylor [Tay86] has many interesting examples of random fractals we cannot even mention in this course, as well as numerous references to earlier works in this field. It is also highly recommended to further study Le Gall's lecture notes [lG92] for a deep analysis of properties of the Brownian path that we only touch upon here.

1.1 Favorite Points and Cover Times

Most Visited Points

Let $S_n = \sum_{i=1}^{n} X_i$ denote a simple random walk in $\mathbb{Z}^2$. A natural question is: "How many times does the walk revisit the most frequently visited site in the first n steps?". This question was posed by Erdős and Taylor more than forty years ago in [ET60] and only recently resolved (in [DPRZ01]). More formally, let $T_n(x)$ denote the number of visits of S_n to x by time n, and let $T_n^* := \max_{x\in\mathbb{Z}^2} T_n(x)$. Then, almost surely,

$$\lim_{n\to\infty} \frac{T_n^*}{(\log n)^2} = \frac{1}{\pi} \tag{1.1}$$

(in Chaps. 4 and 5 we outline the proof of (1.1)). For any $0 < \alpha < 1$ we call $x \in \mathbb{Z}^2$ an α-favorite point if $T_n(x) \geq (\alpha/\pi)(\log n)^2$ and let $\mathcal{F}_n(\alpha)$ denote the set of α-favorite points. It is also proved in [DPRZ01] that for $\alpha \in (0, 1]$,

$$\lim_{n\to\infty} \frac{\log |\mathcal{F}_n(\alpha)|}{\log n} = 1 - \alpha \qquad a.s. \tag{1.2}$$

that is,

$$|\{x : T_n(x) \geq (\alpha/\pi)(\log n)^2\}| \sim n^{1-\alpha} .$$

In other words, the n^{β}-most visited point by time n is visited approximately

$$\frac{1-\beta}{\pi}(\log n)^2$$

times. Moreover, any random sequence $\{x_n\}$ in $\mathbb{Z}^2$ such that $T_n(x_n)/T_n^* \to 1$ must satisfy

$$\lim_{n\to\infty} \frac{\log |x_n|}{\log n} = \frac{1}{2} \qquad a.s.$$

This is a partial reply to a question of Révész (see Chap. 19 of [Rév90]): "What is the rate of convergence of the favorite point of the walk to infinity?" The analogous statement for the simple random walk on $\mathbb{Z}$ is contained in a well-known result of Bass and Griffin [BG85] (see also [LS04] and the references therein for recent developments in settling this question). The proof of these facts tells us also how the most visited site x_n^* is visited: typically the walk makes excursions of "all" time scales n^θ $(0 < \theta < 1)$ between the visits to x_n^*.

Most of the problems we mention can be expressed in a Brownian setting. Indeed, for any Borel measurable function f from $0 \le t \le T$ to $\mathbb{R}^2$, let μ_T^f denote its *occupation measure*:

$$\mu_T^f(A) = \int_0^T \mathbf{1}_A(f_t)dt$$

for all Borel sets $A \subseteq \mathbb{R}^2$. Let $\{w_t\}_{t\ge 0}$ denote the planar Brownian motion started at the origin, and $\bar{\theta} = \inf\{t : |w_t| = 1\}$ the exit time of the unit disc $D(0,1)$ (where throughout $D(x,r)$ denotes the open disc in $\mathbb{R}^2$ of radius r centered at x). Since the path $\{w_t : 0 \le t \le \bar{\theta}\}$ is a compact set, it follows that $\mu_{\bar{\theta}}^w(D(x,r)) = 0$ for any x not in the path and all r small enough. It can be seen (using for example Lévy's uniform modulus of continuity for the upper bound, and techniques as those in [PT87] for the lower bound) that

$$\mathbb{P}\left(\frac{\log \mu_{\bar{\theta}}^w(D(x,r))}{\log r} \underset{r\to 0}{\to} 2 \quad \forall x \in \text{path}\right) = 1\,.$$

Therefore, standard multi-fractal analysis must be refined in order to distinguish between highly visited and less visited points. This was done in [DPRZ01] where it is proved that for any $0 < a \le 2$,

$$\dim\{x : \lim_{r\to 0} \frac{\mu_{\bar{\theta}}^w(D(x,r))}{r^2(\log r)^2} = a\} = 2 - a \qquad a.s. \tag{1.3}$$

(throughout this chapter dim denotes the Hausdorff dimension of the set, see Sect. 3.1 for its definition and some of its properties). In Chap. 5 we outline the proof of (1.3) from a similar result on binary trees derived in Chap. 4. This, together with the appropriate upper bound yields

$$\lim_{r\to 0} \sup_{x\in\mathbb{R}^2} \frac{\mu_{\bar{\theta}}^w(D(x,r))}{r^2(\log r)^2} = 2 \qquad a.s. \tag{1.4}$$

as conjectured by Perkins-Taylor in [PT87]. Note that for a typical x on the Brownian path,

$$\mu_{\theta}^{w}(D(x,r)) \asymp r^2|\log r|$$

(e.g. see [DPRZ01, Lemma 2.1]), so the *a-thick points*, i.e. those in the set considered in (1.3), correspond to unusually large occupation measure. The strong approximation of random walks by Brownian motion relates (1.1) and (1.2) of the discrete setting to (1.4) and (1.3). This derivation highlights the significance of the construction of Komlós-Major-Tsunády [KMT75] which asserts the existence of a Brownian motion w and a simple random walk S on the same probability space such that $|S_{[t]} - w_t| = O(\log t)$. Earlier approximations are not sharp enough for our task (for more details, see Chap. 5).

The proof of (1.3) and (1.4) relies on observing the Brownian motion upon hitting a sequence of concentric discs. When the radii of the discs are appropriately chosen, the observed process is approximately a simple random walk. The authors of [DPRZ01] study the probability of having numerous excursions at many scales (radii) around the same point. During each excursion, the Brownian motion "scores" some occupation measure around that point. Those excursions are independent and since their number is large the total occupation measure is highly concentrated around its mean. An alternative, simpler approach is to discretize the problem by taking a maximal collection of points in $D(0,1)$ such that $\inf_{l\neq j}|x_l - x_j| > \varepsilon$, and consider the random variable

$$Z = \sum_j \mathbf{1}_{\left\{\frac{\mu(D(x_j,r))}{r^2(\log(r))^2} \geq a\right\}}$$

(where we use μ for μ_{θ}^{w}). Indeed, for any point x there exists a j such that x_j is close to x so the occupation measure around x is approximately the same as the one around x_j. Thus, the event $\{Z \geq 1\}$, is approximately the same as

$$\sup_{x\in D(0,1)} \frac{\mu(D(x,r))}{r^2(\log r)^2} \geq a \, .$$

Since it is easily checked that

$$\mathbb{P}\left(\frac{\mu(D(0,r))}{r^2(\log r^2)} > \xi\right) \underset{r\to 0}{\to} e^{-\xi} \, ,$$

one can get upper bounds in (1.3) and (1.4) using the first moment method (i.e. bounding $\mathbb{P}(Z \geq 1)$ by $\mathbb{E}Z$). This is what Perkins and Taylor did in [PT87]. However, due to high correlations between the occupation measure at different points, the second moment of Z is too large, so applying the second moment method fails to produce a tight lower bound.

Late Points and Cover Time

Simulating a simple random walk on a 512×512 torus we observe that those points which are visited late appear as "islands" of various sizes in the simulation. A natural question is what are the geometric characteristics of such islands. Recent results on this topic, motivated by the physics paper [BH91] can be found in [DPRZ05]. More precisely, let $\tilde{X}_j$ be a simple random walk on the two dimensional torus $\mathbb{Z}_n^2 = \mathbb{Z}^2/n\mathbb{Z}^2$, with $\tau_x = \min\{j \geq 0 : \tilde{X}_j = x\}$ the first hitting time of x, and

$$C_n = \max_{x \in \mathbb{Z}_n^2} \{\tau_x\} ,$$

the cover time of $\mathbb{Z}_n^2$ by the simple random walk.

The simple random walk is a time-reversible Markov chain. There already exists a theory dealing with asymptotic for cover time of such processes to which the manuscript [AF01] is devoted. Whereas this theory provides the correct scaling in n it fails to provide the multiplying constant. Indeed, it was only recently shown in [DPRZ04] that

$$\lim_{n\to\infty} \frac{C_n}{(n \log n)^2} = \frac{4}{\pi} \quad \text{in probability} . \tag{1.5}$$

Previous work on this problem include the proof of the upper bound by D. Aldous [Ald89], as well as the proof of the lower bound $2/\pi$ by G. Lawler [Law92]. A nice informal description of this problem, due to H. Wilf [Wil89], is given in the introduction to [AF01]. This problem and those we described in the previous subsection, are much easier to handle in dimension $d \geq 3$. However, little is known beyond the limit of (1.5) or the corresponding limit for $d \geq 3$. For example,

Open problem 1. *Does $\sqrt{C_n}$ − Med$\{\sqrt{C_n}\}$ multiplied by some appropriate normalization factor converge in distribution to a non-degenerate random variable, and if so what are the factor and the limit distribution?* Even the existence of a normalizing sequence that results with a tight, yet non-degenerate collection, is not obvious. For the corresponding problem for simple random walk on regular trees, a proof of tightness is contained in [BZ05].

As shown in [DPRZ04], the strong approximation theorems of [KMT75] allow one to obtain (1.5) from the corresponding problem for the Brownian motion on the unit torus which we describe next. Let $\{X_t\}$ denotes a Brownian motion on the two dimensional unit torus $\mathbb{T}^2$, with the corresponding hitting times,

$$\tau(x, \varepsilon) = \inf\{t > 0 \ : X_t \in D(x, \varepsilon)\} ,$$

and the ε-cover time,

$$C_\varepsilon = \sup_{x \in \mathbb{T}^2} \{\tau(x, \varepsilon)\} . \tag{1.6}$$

Equivalently, C_ε is the amount of time needed for the Wiener sausage of radius ε to completely cover $\mathbb{T}^2$. Then, it is shown in [DPRZ04] that

$$\lim_{\varepsilon\to 0} \frac{C_\varepsilon}{(\log \varepsilon)^2} = \frac{2}{\pi} \qquad \text{a.s.} \tag{1.7}$$

The cover time problem (1.7) is, in a sense, "dual" to the Perkins-Taylor conjecture (1.4), in that it replaces "extremely large" occupation measure by "extremely small" occupation measure.

The general theory of cover times for Markov chains gives universal bounds. For example (see [AF01]), there exist constants k and K such that for any graph G with $|V|$ vertices

$$k|V|\log|V| \le C_G \le K|V|^3 ,$$

where C_G stands for the expected cover time of G by a simple random walk on this graph. In [JS00], Jonasson and Schramm proved that if G is a planar graph of maximal degree d then there are constants k_d and K_d depending only on d such that

$$k_d|V|(\log|V|)^2 \le C_G \le K_d|V|^2 . \tag{1.8}$$

Open problem 2. *Is the square lattice asymptotically the easiest to cover when $d=4$? That is, does $C_G \ge (1/\pi)|V|(\log|V|)^2(1+o(1))$ as $|V|\to\infty$, for any collection of planar graphs of maximal degree $d=4$?*

1.2 Fractal Geometry of Late and Favorite Points

Returning to the Brownian motion on the two dimensional unit torus $\mathbb{T}^2$, en route to (1.7) it is shown in [DPRZ04] that

$$\sup_{x\in\mathbb{T}^2}\limsup_{\varepsilon\to 0}\frac{\tau(x,\varepsilon)}{(\log\varepsilon)^2}=\frac{2}{\pi}\quad a.s. ,$$

and that for $a\le 2$,

$$\dim\left\{x\in\mathbb{T}^2:\ \limsup_{\varepsilon\to 0}\frac{\tau(x,\varepsilon)}{(\log\varepsilon)^2}=\frac{a}{\pi}\right\}=2-a\quad a.s. \tag{1.9}$$

We call $x\in\mathbb{T}^2$ a *late point* if it is in the set considered in (1.9) for some $a>0$.

Open problem 3. *Study the consistently late points, where the* lim sup *in (1.9) is replaced by a limit or a* lim inf. The difficulty in doing so lies in the fact that the behaviors for different scales (i.e. ε's) are highly dependent when the scales are too close to each other.

Moving back to the discrete setting of a simple random walk (SRW) on the lattice torus $\mathbb{Z}_n^2$ which starts at the origin, the corresponding set of α-late points is now

$$\mathcal{L}_n(\alpha)=\{x\in\mathbb{Z}_n^2:\ \tau_x\ge\alpha(4/\pi)(n\log n)^2\} .$$

It is shown in [DPRZ05] that for $\alpha\in(0,1]$