
Volume 4, Issue 2, May 2010
ISSN 1574-0137
ELSEVIER
Computer Science Review
Editors-in-Chief:
Josep Díaz
Jaroslav Nešetřil

COMPUTER SCIENCE REVIEW 4 (2010) 125–127

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cosrev

Book review

Devdatt P. Dubhashi, Alessandro Panconesi (Eds.), Concentration of Measure for the Analysis of Randomized Algorithms. Cambridge University Press, UK (2009). ISBN: 978-0-521-88427-3

1. Introduction

The phenomenon of *concentration of measure* was initially studied by mathematicians in the framework of measure theory and probability theory. At the end of the 19 century, Emile Borel, Andrey Markov and Pafnuty Chebyshev gave probabilistic interpretations of the concentration of measure phenomenon for single and multiple independent random variables. Previously, asymptotic "concentration" results were known, such as the *Central Limit Theorem* and the *de Moivre–Laplace Theorem* for the sum of independent random variables under certain conditions, results which lie at the core of the concentration of measure phenomena.

In very rough words, in probability theory the concentration of measure states that given n random variables, each with $O(1)$ mean and variance, and all them independent of the others (or almost), their linear combination is concentrated around their mean. In fact, as we will mention, the concentration of measure phenomenon also occurs for non-linear functions of the random variables, which moreover do not need to be independent, under certain conditions of the function and variables. The book by Michael Ledoux is one of the most general tracts written on the topic [1].

In the last decades, due to the increasing use of randomized algorithms, the analytic techniques used to prove the concentration of measures have become an important tool in the fields of *computer science* and *discrete mathematics*. For instance, when performing the average analysis of a deterministic algorithm for a problem, for an instance chosen randomly from a given distribution of the instances, we define the running time of the algorithm as a random variable and, in addition to computing the expected value of the random variable, we wish to prove that the distribution of the random variable is sharply concentrated around the expectation, in order to avoid the *lottery paradox*, when a few biased instances could significantly deviate the value of the expectation.

Concentration of measure for the analysis of randomized algorithms by Devdatt Dubhashi and Alessandro Panconessi, Cambridge UP, 2009, comes as the first book to deal fully with the topic of concentration inequalities, which is targeted specifically to the computer science and discrete mathematics community. The previously mentioned books by [1] and [2], are written from a measure theoretical point of view, which could represent a drawback for a part of the computer science community. However, several books written in the last decade have excellent chapters and sections on the topic of concentration bounds, for example [3–7], and also there have been a few general surveys on concentration bounds specifically addressed to researchers in the fields of CS and discrete mathematics, among others [8–13]. The reader can find a more exhaustive list in the introduction section of the last citation. Recently, a nice short introduction to the topic has appeared in Terence Tao's blog *What's new* on January 3rd. 2010 [14]).

The authors of the reviewed book, Dubhashi and Panconessi, are two well-known researchers, with plenty of nice papers using a variety of concentration bounds. Two very qualified people to write a book on this topic, which is addressed to the computer science and discrete math community. The book introduces the basic theory for each of the different techniques, and uses examples to show how to apply each technique. Personally, I like books that do not rely too much on leaving the non-trivial proofs of important theorems as exercises for the reader, even if the book gives pointers to places where the proofs could be found, in particular when dealing with the basic material. The book assumes a *certain* degree of maturity in probability theory: for instance Boole's union-bound is often used but never defined; In Chapter 1 there is a very brief statement about why concentration bounds are important, but unless you already know the problem, it is not sufficient; Markov's inequality appears mentioned in the derivation of Chernoff bound, but it is never mentioned that for many basic applications, Markov inequality suffices. Personally I believe than a book should also offer an early historical perspective of the field. Therefore, as the authors mention in the introduction, the book is aimed at people that already know the basic techniques of probability, for example at the level of the first four chapters in [4].

1574-0137/$ - see front matter
doi:10.1016/j.cosrev.2010.03.003

2. Chapter-by-chapter synopsis and analysis

The first chapter deals with the Chernoff–Hoeffding bounds. Starting from the moment generating function of a random variable, the authors derive the basic Chernoff bound for the linear combination of independent binomial random variables. Afterwards the authors present the Hoeffding extension and several useful versions of the Chernoff bounds, including Berstaein's inequality, the proof of which is left as an exercise. The exercises provide further versions of the Chernoff theorems applied to different settings of independent random variables. Chapter 2 shows how to apply these results to concrete problems, such as probabilistic amplification, load balancing, skip lists, quicksort and a delicious example; a short proof of the Jhonson–Lindenstrauss theorem, on the embedding of *n* points in an Euclidean real space into a lower dimension Euclidean space, such that the embedding has low distortion.

Chapter 3 deals with Chernoff–Hoeffding bounds for applications with limited independence. It is a very nice collection of results and their application to examples. The first section deals with Chernoff–Hoeffding bounds for negative dependent random variables with applications to several problems (one of the authors of the book has a very nice paper on this topic). Afterwards, they deal with some applications of Janson's inequality, a Chernoff–Hoeffding bound for random variables with limited dependency and a bound for a sequence of random variables generated by a Markov chain under certain conditions. The authors apply the last bound to obtain an algorithm for probability amplification.

In Chapter 4, the authors present Karp's Theorems on probabilistic recurrences, where the iterating term is a random variable. They state the bounds without proofs and show their use in four examples.

In Chapter 5 they state and prove Azuma's inequality and derive the simple and the average bounded difference methods. In the next two chapters they present an array of examples to show the applicability of the two *bounded difference* methods. Among the examples one of the authors' favorite; the *distributed edge coloring*, refinements of which are also presented in the chapters that follow.

Chapters 8 and 9 present alternatives to the average bounded difference method, when the function under consideration is not Lipschitz (but has some nice "average" properties). The method of average bounded variance is explained quite rigorously in Chapter 8. Chapter 9 sketches two further particular techniques for non Lipschitz functions: The Kim-Vu multivariate polynomial inequality and the Janson–Rucinski's deletion method.

Isoperimetric inequalities can be considered as concentration results for functions which go asymptotically fast to 0. In Chapters 10 to 14 the authors explore this topic of research, which has been important in the mathematical literature, see for example [1]. In Chapter 10, the authors introduce the topic and relate some of the previous concentration bounds (Chernoff's inequality or the method of bounded differences) to some isoperimetric inequalities. One of the early isoperimetric inequalities than the authors mention and use is Harper's inequality for the Hamming cube. It is a pity that the authors do not cite as reference the original paper by the author or better the nice book by Harper on isoperimetric problems [15]. In Chapter 11, Talagran's inequality is stated, which is a very powerful tool that can be explained in an algorithmic way [5,3] (see also the nice exposition of Talagrand in Terence Tao's blog: *What's new*, June the 9th 2009 [16]). The method is used to find sharp concentrations for several problems: Random minimum spanning tree, the traveling salesman problem in a Euclidean graph, edge coloring. Chapter 12 gives a succinct introduction to transportation cost inequalities in product spaces and the isoperimetric inequalities that can be derived with this technique. Chapter 13 proves a general version of Talagran's inequality, using the transportation cost method.

Chapter 15 gives a sketchy introduction to log-Sobolev inequalities and their use in deriving concentration of measure results. The technique is used to derive new proofs for the method of bounded differences and Talagran's inequality. The book finishes with a *"cheat sheet"* on the most useful bounds described in the book. For some reason, the authors left Talagrand's inequality outside it, which, as the authors mention in Chapter 11, Talagran's inequality could be a useful tool when dealing with some kinds of problems, like geometric graphs' problems. May be the omission is due to the fact that the presentation of Talagran's inequality is too general, and maybe the version given in [5] is more friendly to be used as a recipe for *the working computer scientist* doing analysis of algorithms.

3. Conclusions

I believe that the book *Concentration of Measure for the Analysis of Randomized Algorithms* is a handy book to have around, especially by researchers working in algorithmics and discrete mathematics. The book is written by two people with a lot of expertise and experience in the topic. As the authors say, the book aims to be a *"cook-book"* to supply recipes to do formal analysis of concentration bounds under different conditions of the random variables. In that sense, through Chapter 8 and also a bit Chapter 11, the objective is accomplished. The last chapters give an introduction to transportation inequalities and log-Sobolev inequalities, the presentation is sufficient to give an idea of the possibilities of the material, and for the interested reader there are good references to a more formal exposition, like [1] and [13]. From this reviewer's point of view, a missing "basic" concentration technique is the *Differential Equation Method* (DEM). The DEM has become an easy to use, nice method to obtain sharp concentrations of random variables, that arise in some kinds of random processes. One of the first applications to a particular discrete process was given in [17]. Wormald gave the general theorem of application and a nice survey [18]. I believe that the basic recipe to use the DEM in several situations could have been an interesting chapter in the book under review, and it could have been introduced with little effort after defining martingales.

As already mentioned, throughout the book there are *exercises* and *problems*. The exercises in the middle of the chapters to complete technical details or proofs, and the problems at the end of each chapter. To my personal taste, even in its style of "*cook-book*" there is an excess of important material left as exercises, some of them far from easy. The problems at the end of the chapters have a very large variance in the degree of difficulty. A nice characteristic of the book is that at the end of each chapter there is a section with references to where the presented material appeared. Through those references we can find the solutions to some of the problems.

REFERENCES

[1] M. Ledoux, The Concentration of Measure Phenomenon, American Mathematical Society, Providence, RI, 2001.

[2] F. den Hollander, Large Deviations, vol. 14, American Mathematical Society, Philadelphia, 2000.

[3] N. Alon, J.H. Spencer, The Probabilistic Method, third ed., Wiley-Interscience, New York, 2008.

[4] M. Mitzenmacher, E. Upfal, Probability and Computing, Cambridge U.P., 2005.

[5] M. Molloy, B. Reed, Graph Coloring and the Probabilistic Method, Springer, 2000.

[6] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.

[7] S. Janson, T. Łuczak, A. Rucinski, Random Graphs, Wiley-Interscience, New York, 2000.

[8] C. McDiarmid, On the method of bounded differences, in: J. Siemons (Ed.), Surveys in Combinatorics, 1989, in: London Mathematical Society Lectures Notes Series, 1989, pp. 669–188.

[9] C. McDiarmid, Concentration, in: H. Habib, C. McDiarmid, J. Ramirez, B. Reed (Eds.), Probabilistic Methods for Algorithmic Discrete Mathematics, Springer-Verlag, 1998, pp. 195–248.

[10] T. Hagerup, C. Rub, A guided tour of Chernoff bounds, Information Processing Letters 33 (1990) 305–308.

[11] J. Díaz, J. Petit, M. Serna, A guide to concentration bounds, in: S. Rajasekaran, P. Pardalos, J. Reif, J. Rolim (Eds.), in: Handbook on Randomized Computing, vol. II, Kluwer Press, New York, 2001, pp. 457–507 (Chapter 12).

[12] D. Dubhashi, Martingales and locality in distributed computing, in: J.R.M. Luby, M. Serna (Eds.), Randomization and Approximation Techniques in Computer Science, in: Lecture Notes in Computer Science, vol. 1518, Springer-Verlag, Berlin, 1998, pp. 60–70.

[13] O. Boucheron, S. Boucheron, G. Lugosi, Concentration inequalities, in: O. Bousquet, U.v. Luxburg, G. Rotsch (Eds.), Advanced Lectures in Machine Learning, Springer, 2004, pp. 208–240.

[14] Terence Tao, Talagran's concentration inequality. What's new 06/09/2009. http://terrytao.wordpress.com//2010/01/03/254a-notes-1-concentration-of-measure/.

[15] L.H. Harper, Global Methods for Combinatorial Isoperimetric Problems, Cambridge University Press, Cambridge, 2004.

[16] Terence Tao, Talagran's concentration inequality. What's new 06/09/2009. http://terrytao.wordpress.com/2009/06/09/talagrands-concentration-inequality/.

[17] T. Kurtz, Solutions of ordinary differential equations as limits of pure Markov jump processes, Journal of Applied Probability 7 (1970) 49–58.

[18] N.C. Wormald, The differential equation method for random graph processes and greedy algorithms, in: M. Karoński, H. Prömel (Eds.), in: Lectures on Approximation and Randomized Algorithms, PWN, Warsaw, 1999, pp. 73–155.

Josep Díaz

Department LSI, Grup de Recerca ALBCOM, UPC Barcelona, Spain

E-mail address: diaz@lsi.upc.edu.