MR2032426 (2004m:60004) 60-02 05-02 05A05 05A16 05D40 60C05 60F17
Arratia, Richard (1-SCA);
Barbour, A. D. [Barbour, Andrew D.] (CH-ZRCH-AM); **Tavaré, Simon** (1-SCA)
★**Logarithmic combinatorial structures: a probabilistic approach.**
EMS Monographs in Mathematics.
European Mathematical Society (EMS), Zürich, 2003. *xii*+363 *pp. ISBN* 3-03719-000-0

This book is devoted to problems arising in the asymptotic enumeration of decomposable combinatorial structures, i.e. structures that can be naturally decomposed into component elements. The usual probabilistic interpretation of such problems assumes a distribution on the set of all such structures of size n is given. A frequently asked question of theoretical and practical importance is to characterize properties of "typical" structures of size n as n becomes large. Here "typical" is understood to mean "with high probability" (or "with probability tending to 1 as $n\to\infty$").

In the present book the authors are specifically concerned with the so-called component frequency spectrum of decomposable structures—that is, with the numbers and sizes of the components of such structures. More precisely, the component spectrum of a decomposable structure of size n consists of numbers $C_i^{(n)},\ i=1,2,\dots,n$, counting components of size i. The discrete non-negative random process $C^{(n)}=(C_1^{(n)},C_2^{(n)},\dots,C_n^{(n)}),\ n=1,2,\dots,$ has dependent coordinates since $\sum_{i=1}^n iC_i^{(n)}=n$. The authors present a comprehensive account of asymptotic results for a large subclass of combinatorial structures—the so-called logarithmic combinatorial structures. Two main assumptions are adapted for those structures: "the conditioning relation" and "the logarithmic condition". The authors' unifying approach to problems concerning limiting distributions, asymptotics of characteristics and rates of convergence of functionals related to $C^{(n)}$ is essentially based on a relation between the joint distribution $\mathcal{L}(C_1^{(n)},\dots,C_n^{(n)})$ and the conditional distribution $\mathcal{L}(Z_1,\dots,Z_n\mid T_{0n}=n)$, where $Z_1,Z_2,\dots$ is a fixed sequence of independent random variables taking values in $\{0,1,\dots\}$ and $T_{0n}=T_{0n}(Z)=\sum_{i=1}^n iZ_i$. The conditioning relation is then given by $\mathcal{L}(C_1^{(n)},\dots,C_n^{(n)})=\mathcal{L}(Z_1,\dots,Z_n\mid T_{0n}=n)$. Furthermore, the logarithmic condition provides the sequence $\{Z_i,i\ge 1\}$ with the following two limits: $iP(Z_i=1)\to\theta,\ iEZ_i\to\theta$, as $i\to\infty$, where $\theta>0$.

The joint distribution of the component counts $C_i^{(n)}$ classifies combinatorial structures into one of the following three major classes: the assemblies, multisets and selections. The probabilistic approach based on the conditioning relation allows one to identify the structures by specifying the distribution of the random variables Z_i. For the classical combinatorial structures (permutations, mappings, set and integer partitions, labeled and unlabeled forests, polynomials over $\mathrm{GF}(q)$, etc.) the Z_i have either Poisson, negative binomial or binomial distributions. In the probabilistic setting adopted in the book, there is, however, no need to specify the distributions of Z_i. This flexibility allows the authors to obtain, with a unified theory, general results for a large variety of combinatorial structures including, for instance, θ-biased random permutations defined by the Ewens sampling formula and polynomials over a finite field.

Asymptotic properties of assemblies, multisets and selections have been studied by many authors without regard to the conditioning relation, but using generating functions and approximations in the Cauchy coefficient formula instead. The book under review presents in contrast the authors' own probabilistic method of study developed in the last two decades. A significant advantage of their study is the possibility of applying various classical and modern probabilistic tools and techniques. Thus, the authors employ probability metrization, functional limit theorems and approximations for discrete and continuous time processes, large deviation estimates, size biasing and Stein's method. Furthermore, the probabilistic approach allows them to achieve a high level of accuracy in their approximations, including bounds on the error terms and large deviation probabilities. It seems to be very hard to study this kind of problems following the traditional method of generating functions and complex integration.

The book is written and organized in an excellent way. Its preface contains a clear description of the authors' goals and methods of proof. A summary of historical remarks is also included there. The material is divided into 14 chapters. The bibliography contains 228 references. As the authors claim, the material can be thematically divided into two parts. The first one is an introduction to the asymptotic theory of decomposable combinatorial structures, while the second one presents a detailed study of this subject in the style of a research monograph.

Chapter 1 summarizes classical and important facts concerning the cycle decomposition of a permutation of n letters and the prime factorization of an integer chosen uniformly at random from the set $\{1, 2, \ldots, n\}$. The authors claim that these two instances are "the oldest and most significant examples of logarithmic structures". In order to illustrate phenomena of further interest related to the joint behavior of component counts of general logarithmic structures, they present a lot of exact and asymptotic results, including central and local limit theorems for the total number of cycles, the behavior of small and large cycles, the Erdős-Turán law of the permutation order, and Feller's coupling method. Then, they outline the analogous results for the prime factorization of a random integer, showing the similarities and contrasts with the cycle decomposition of permutations.

The aim of Chapter 2 is to introduce and describe decomposable combinatorial structures, both logarithmic and non-logarithmic. The authors start this by considering some typical examples. They include permutations, mappings from an n-element set into itself, mapping patterns, forests of labeled and unlabeled trees, polynomials over $\mathrm{GF}(q)$, etc. This description continues further in terms of the general classes: assemblies, multisets and selections. The authors give a probabilistic interpretation of these combinatorial objects, focusing on the conditioning relation and the logarithmic condition. The chapter ends with probabilistic comments on refining and coloring. An extension of coloring, called tilting, is also considered there.

Chapter 3 focuses on logarithmic structures satisfying the conditioning relation without specific restrictions on the sequence of independent non-negative integer-valued random variables Z_i. All classical examples considered previously are included in this model. The chapter deals with the accuracy of the probabilistic approximations for such structures. The authors assess it using probability metrics—total variation distance and various Wasserstein distances. The chapter concludes with an introduction to Stein's method.

Chapters 4 and 5 are devoted to the family of discrete probability distributions obtained by the Ewens sampling formula. In Chapter 4 the method of size-biasing is discussed in order to be subsequently applied to obtain local limit approximations for $T_{0n} = T_{0n}(Z)$ and $T_{bn} = \sum_{j=b+1}^{n} jZ_j$, when $b, n \to \infty$, so that $b/n \to \alpha \in (0, 1]$. The Ewens sampling formula also has a meaningful combinatorial interpretation: it defines the joint distribution of the component counts for a permutation of n letters, chosen with probability biased by $\theta^{K_{0n}}$, where K_{0n} is the number of cycles and $\theta > 0$. The choice $\theta = 1$ implies the classical formula for uniform random permutations. θ-biased permutations defined in this way satisfy the conditioning relation with independent random variables Z_i having Poisson distributions with parameters equal to θ/i. The main goal of Chapter 5 is to show that properties of uniform random permutations discussed in Chapter 1 carry over almost unchanged to θ-biased random permutations. The authors focus on asymptotic properties of such permutations and discuss in detail the limiting processes that are found upon passage to the limit—the scale invariant Poisson process, the GEM distribution, and the Poisson-Dirichlet distribution.

In Chapter 6 the authors adapt the methods applied to the Ewens sampling formula to a more general class of logarithmic combinatorial structures that satisfy the conditioning relation with non-negative and integer-valued Z_i's, which are not necessarily Poisson distributed, as in the previous chapter. Here the main results extend the limit theorems for small and large components of such logarithmic structures.

Chapter 7 is devoted to structures that are more general than assemblies, multisets and selections. Their study needs more sophisticated tools. The authors use the conditioning relation to show that the joint distribution of the large components of a general logarithmic combinatorial structure is close to that of the large components in the Ewens sampling formula, provided that, for large i, the distribution of the Z_i's is close to that of certain Z_i^*'s, which have Poisson distribution with parameters $\theta/i,\ \theta > 0$, and the distribution of $T_{0n}(Z)$ is close to that of $T_{0n}(Z^*)$ for large n. As a first step, they establish working conditions under which the differences between $\mathcal{L}(Z_i)$ and $\mathcal{L}(Z_i^*)$ are visible and controllable. Then, using these conditions, the authors apply Stein's method to show the closeness between $\mathcal{L}(T_{0n}(Z))$ and $\mathcal{L}(T_{0n}(Z^*))$, and after that, between $\mathcal{L}(C_1^{(n)}, \dots, C_b^{(n)})$ and $\mathcal{L}(Z_1, \dots, Z_b)$ for $b = o(n)$. In this way, they obtain a result for small components. The illustrations of the method include random mappings and random polynomials over a finite field. To proceed with large components the authors refine some limit theorems from Chapter 3 by giving error bounds in both cases of small and large components. Local and global approximations are stated there. The proofs of all of the facts and theorems of this theory are postponed to the technical part of the book—Chapters 9 through 13.

Chapter 8 contains a number of consequences of the approximation theorems of the preceding chapter. The authors illustrate the power of their method by considering functional limit theorems (Brownian motion asymptotics), Poisson-Dirichlet limits, asymptotics for the number of components and the Erdős-Turán law. Thus, earlier limiting results are improved in two ways. First, the authors extend their validity to a very general class of general logarithmic combinatorial structures, and second, they assess errors in the distributional approximations under appropriate probability metrics.

The last chapter, Chapter 14, also belongs to the technical part of the book. It collects together some technical lemmas that are used without proofs elsewhere in the text.

The book is clearly intended as a textbook for graduate students and researchers interested in probabilistic combinatorics. The authors succeed in presenting their powerful method for logarithmic combinatorial structures in a clear and rigorous way. The variety of illustrative examples that are used throughout the whole text make their approach much more accessible. The book would be also an ideal and comprehensive resource for mathematicians working in related areas, such as enumerative combinatorics, probabilistic limit theorems and analysis of algorithms.

Lyuben R. Mutafchiev