Progress in Probability
69

Ramsés H. Mena
Juan Carlos Pardo
Víctor Rivero
Gerónimo Uribe Bravo
Editors

XI Symposium on Probability and Stochastic Processes

CIMAT, Mexico, November 18-22, 2013

Birkhäuser

Birkhäuser

Progress in Probability

Volume 69

More information about this series at
http://www.springer.com/series/4839

Ramsés H. Mena • Juan Carlos Pardo •
Víctor Rivero • Gerónimo Uribe Bravo
Editors

XI Symposium on Probability and Stochastic Processes

CIMAT, Mexico, November 18-22, 2013

Birkhäuser

Editors
Ramsés H. Mena
Probabilidad y Estadística
IIMAS UNAM
Mexico

Juan Carlos Pardo
CIMAT
Guanajuato, Mexico

Víctor Rivero
CIMAT
Guanajuato, Mexico

Gerónimo Uribe Bravo
Instituto de Matemáticas
UNAM
Mexico

ISSN 1050-6977 ISSN 2297-0428 (electronic)
Progress in Probability
ISBN 978-3-319-13983-8 ISBN 978-3-319-13984-5 (eBook)
DOI 10.1007/978-3-319-13984-5

Library of Congress Control Number: 2015945576

Mathematics Subject Classification (2010): 60-06

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.birkhauser-science.com)

Introduction

The present volume contains contributions and lecture notes of the XI Symposium on Probability and Stochastic Processes, held at Centro de Investigación en Matemáticas, México (CIMAT), November 18–22 in 2013.

This Symposium traces its roots back to December 1988 at CIMAT, when it was held for the first time, and constitutes one of the main events in the area happening biannually in various academic institutions in Mexico. During these more than 25 years, this series of symposia continuously accomplishes its main goal of exchanging ideas and discussing the current developments in the field by gathering national and international researchers as well as graduate students.

On this occasion, the Symposium was part of many activities organised in Mexico to celebrate the International Year of Statistics 2013. It gathered scholars from over 10 countries and included a wide set of topics that highlighted the interaction between statistics and stochastic processes. The scientific programme included two courses: 'Probabilistic aspects of minimal spanning trees' by Louigi Addario-Berry and 'Spatial point pattern analysis' by Carlos Díaz-Avalos. The event also benefited from six plenary conferences delivered by Loïc Chaumont, Janos Englander, Enrique Figueroa, Daniel Hernández, Andreas Kyprianou and Mark Podolskij; eight thematic sessions; eight contributed talks; and several poster presentations.

The volume begins with the lecture notes by Addario-Berry, providing with an accessible description of some features of the multiplicative coalescent and its connection with random graphs and minimum spanning tree. The tutorial is then followed by the illustrative article by Arizmendi and Gaxiola where they show that the large N-fold limit of the spectral distribution of a connected graph converges to a certain centred Bernoulli distribution. Baudoin studies stochastic differential equations driven by Brownian loops in a free Carnot group providing with sufficient conditions to ensure that solutions admit smooth densities. By considering a simple age distribution, Blath, Eldon, González Casanova and Kurt investigate the behaviour of the genealogy of a Wright-Fisher population model under the influence of a strong seed-bank effect. A Vervaat-like pathwise construction of a process with cyclically exchangeable increments with a predetermined minimum is introduced and analysed by Chaumont and Uribe; it consists of inverting the paths of the process at an adequately chosen random time τ, which surprisingly happens to be independent and to follow a uniform distribution. Motivated by different extensions of the so-called Bercovici-Pata bijection, Dominguez-Molina and Rocha Arteaga obtain an interesting sample path representation by covariation processes of matrix Lévy processes of rank-one jumps. Using the thermodynamic interpretation of the α parameter of a Poisson-Dirichlet distribution, i.e. as the ratio between the temperature T and a critical temperature Tc, Feng and Zhou study the asymptotic behaviour of such distribution as the temperature approaches the critical value. Gordienko, Martinez and Ruiz de Chavez provide estimates of the stability index with respect to the total variation metric and Prokhorov distance for a total-reward Markov decision chains with an absorbing set. The problem of determining a price for a contingent claim in an incomplete market is analysed by Hernández-Hernández and Sheu; they propose a pricing based on the utility, which is assumed to be exponential, and show that there exists a unique solution to optimal control problems. Relying heavily on the use of the so-called Dynkin-Kuznetsov N-measures, Murillo and Pérez Garmendia provide a pathwise backbone decomposition for supercritical superprocesses with nonlocal branching. Their result complements a related result obtained for supercritical superprocesses without nonlocal branching. Pedersen and Sauri present a detailed study of the stationary distribution of Lévy semi-stationary processes with a gamma kernel; they establish conditions for absolute continuity, infinite divisibility and self-decomposability, together with descriptions of its characteristics. The paper by Podolskij constitutes a thorough review of the theory of ambit fields, which is a flexible model for dynamical structures in time and/or space; the interesting list of open problems included will challenge any reader and generate deep research in the field.

In summary, the high quality and variety of these contributions give a broad panorama of the rich academic programme of the Symposium and of its impact. It is worth noting that all papers, including the invited course lecture notes, were subject to a strict refereeing process with high international standards. We are very grateful to the referees, many of whom are leading experts in their fields, for their careful and useful reports. Their comments were implemented by the authors allowing to improve the material here presented.

We would also like to extend our gratitude to all the authors whose original contributions appear published here as well as to all the speakers and session organisers in the Symposium for their stimulating talks and support. Their valuable contributions encourage the interest and activity in the area of probability and stochastic processes in Mexico.

We hold in high regard the editors of the series Progress in Probability: Davar Khoshnevisan, Andreas E. Kyprianou and Sidney I. Resnick for giving us the possibility to publish the Symposium Volume in this prestigious series.

Special thanks go to the Symposium venue CIMAT and its staff for its great hospitality and for providing excellent conference facilities. We are indebted to Rosy Davalos whose outstanding organisational work permitted us to concentrate mainly in the academic aspects of the conference.

The Symposium as well as the edition of this volume would not have been possible without the generous support of our sponsors: Centro de Investigación en Matemáticas, International Year of Statistics 2013 (Mexico), Laboratorio Internacional Solomon Lefschetz CNRS-CONACYT, Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas and Instituto de Matemáticas, UNAM.

Finally, we hope the reader of this volume enjoys learning about the various topics treated as much as we did editing it.

Ramsés H. Mena Chavez
Juan Carlos Pardo
Víctor Manuel Rivero
Gerónimo Uribe

Previous Volumes from the Symposium on Probability and Stochastic Processes

- M. E. Caballero and L. G. Gorostiza, editors. *Simposio de Probabilidad y Procesos Estocásticos*, volume 4 of *Aportaciones Matemáticas: Notas de Investigación [Mathematical Contributions: Research Notes]*. Sociedad Matemática Mexicana, México, 1989.

  Held at the Universidad de Guanajuato, Guanajuato, December 12–14, 1988.
- M. E. Caballero and L. G. Gorostiza, editors. *II Simposio de Probabilidad y Procesos Estocásticos. I Encuentro México-Chile de Análisis Estocástico*, volume 7 of *Aportaciones Matemáticas: Notas de Investigación [Mathematical Contributions: Research Notes]*. Sociedad Matemática Mexicana, México, 1992.

  Held at the Universidad de Guanajuato, Guanajuato, January 27–31, 1992.
- M. E. Caballero and L. G. Gorostiza, editors. *III Simposio de Probabilidad y Procesos Estocásticos*, volume 11 of *Aportaciones Matemáticas: Notas de Investigación [Mathematical Contributions: Research Notes]*. Sociedad Matemática Mexicana, México, 1994.

  Held at the Universidad de Sonora, Hermosillo, March 7–11, 1994.
- L. G. Gorostiza, J. A. León, and J. A. López-Mimbela, editors. *IV Simposio de Probabilidad y Procesos Estocásticos*, volume 12 of *Aportaciones Matemáticas: Notas de Investigación [Mathematical Contributions: Research Notes]*. Sociedad Matemática Mexicana, México, 1996.

  Held at the Centro de Investigación en Matemáticas, Guanajuato, April 1–5, 1996.
- J. M. González Barrios and L. G. Gorostiza, editors. *Modelos estocásticos*, volume 14 of *Aportaciones Matemáticas: Investigación [Mathematical Contributions: Research]*. Sociedad Matemática Mexicana, México, 1998. Papers from the 5th Symposium on Probability and Stochastic Processes.

  Held at the Centro de Investigación en Matemáticas, Guanajuato, March 29–April 3, 1998.
- D. Hernández, J. A. López-Mimbela, and R. Quezada, editors. *Modelos estocásticos II*, volume 16 of *Aportaciones Matemáticas: Investigación [Mathematical Contributions: Research]*. Sociedad Matemática Mexicana, México, 2001. Papers from the 6th Symposium on Probability and Stochastic Processes.

  Held at the Centro de Investigación en Matemáticas, Guanajuato, May 23–27, 2000.
- J. M. González-Barrios, J. A. León, and A. Meda, editors. *Stochastic models*, volume 336 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI; Sociedad Matemática Mexicana, México, 2003. Aportaciones Matemáticas. [Mathematical Contributions]. Proceedings of the 7th Symposium on Probability and Stochastic Processes.

  Held at the National University of Mexico, Mexico City, June 23–28, 2002

- M. Bladt, J. A. Lopez Mimbela, J. Ruiz de Chavez, and R. Navarro, editors. *Special issue: Proceedings of the 8th Symposium on Probability and Stochastic Processes*, volume 22. Stochastic Models, 2006.

  Held at the Universidad de las Americas in Puebla, June 20–25, 2004.
- M. E. Caballero, V. Rivero, and J. Ruiz de Chávez, editors. *Ninth Symposium on Probability and Stochastic Processes*, volume 24. *Stochastic Models*, 2008.

  Held at the Centro de Investigación en Matemáticas, Guanajuato, November 20–24, 2006.
- M. E. Caballero, L. Chaumont, D. Hernández-Hernández, and V. Rivero, editors. *X Symposium on Probability and Stochastic Processes and the First Joint Meeting France-Mexico of Probability*, volume 31 of *ESAIM Proceedings*. EDP Sciences, Les Ulis, 2011.

  Held at the Centro de Investigación en Matemáticas, Guanajuato, November 3–7, 2008.

Contents

Partition Functions of Discrete Coalescents: From Cayley's Formula to Frieze's $\zeta(3)$ Limit Theorem

Louigi Addario-Berry

Abstract In these expository notes, we describe some features of the multiplicative coalescent and its connection with random graphs and minimum spanning trees. We use Pitman's proof (Pitman, J Combin Theory Ser A 85:165–193, 1999) of Cayley's formula, which proceeds via a calculation of the partition function of the additive coalescent, as motivation and as a launchpad. We define a random variable which may reasonably be called the empirical partition function of the multiplicative coalescent, and show that its typical value is exponentially smaller than its expected value. Our arguments lead us to an analysis of the susceptibility of the Erdős-Rényi random graph process, and thence to a novel proof of Frieze's $\zeta(3)$-limit theorem for the weight of a random minimum spanning tree.

1 Introduction

Consider a discrete time process $(P_i, 1 \leq i \leq n)$ of coalescing blocks, with the following dynamics. The process starts from the partition of $[n] = \{1, \ldots, n\}$ into singletons: $P_1 = \{\{1\}, \ldots, \{n\}\}$. To form P_{i+1} from P_i choose two parts P, P' from P_i and merge them. We assume there is a function κ such that the probability of choosing parts P, P' is proportional to $\kappa(|P|, |P'|)$; call κ a *rate kernel.*

Different rate kernels lead to different dynamics. Three kernels whose dynamics have been studied in detail are $\kappa(x, y) = 1$, $\kappa(x, y) = x + y$, and $\kappa(x, y) = xy$; these are often called Kingman's coalescent, the additive coalescent, and the multiplicative coalescent, respectively. In these cases there is a natural way to enrich the process and obtain a *forest-valued* coalescent.

These notes are primarily focussed on the properties of the forest-valued multiplicative coalescent. We proceed from a statistical physics perspective, and begin by analyzing the partition functions of the three coalescents. Here is what we mean by this. Say that a sequence $(P_1, \ldots, P_n)$ of partitions of $[n]$ is an *n-chain* if

L. Addario-Berry (✉)
Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, QC H3A 2K6, Canada
e-mail: louigi@math.mcgill.ca

R.H. Mena et al. (eds.), *XI Symposium on Probability and Stochastic Processes*, Progress in Probability 69, DOI 10.1007/978-3-319-13984-5_1

$P_1 = \{\{1\}, \ldots, \{n\}\}$ is the partition of n into singletons, and for $1 \le i < n$, P_{i+1} can be formed from P_i by merging two parts of P_i. Think of $\kappa(x, y)$ as the number of possible ways to merge a block of size x with one of size y. Then corresponding to an n-chain $P = (P_1, \ldots, P_n)$ there are

$$\prod_{i=1}^{n-1} \kappa(|A_i(P)|, |B_i(P)|)$$

possible ways that the coalescent may have unfolded; here we write $A_i(P)$ and $B_i(P)$ for the blocks of P_i that are merged in P_{i+1}. Writing $\mathcal{P}_n$ for the set of n-chains, it follows that the total number of possibilities for the coalescent with rate kernel κ is

$$\sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \prod_{i=1}^{n-1} \kappa\left(|A_i(P)|, |B_i(P)|\right) ,$$

and we view this quantity as the partition function of the coalescent with kernel κ.

The partition functions of Kingman's coalescent and the additive and multiplicative coalescents have particularly simple forms: they are

$$\begin{aligned} Z_{\mathrm{KC}}(n) &= n!(n-1)!\,, \\ Z_{\mathrm{AC}}(n) &= n^{n-1}(n-1)!\,, \text{ and} \\ Z_{\mathrm{MC}}(n) &= n^{n-2}(n-1)!\,. \end{aligned}$$

These formulae are proved in Sect. 2. A corollary of the formula for $Z_{\mathrm{KC}}(n)$ is that the number of increasing trees with n vertices is $(n-1)!$; this easy fact is well-known. The formula for $Z_{\mathrm{AC}}(n)$ is due to Pitman [13], who used it to give a beautiful proof of Cayley's formula; this is further detailed in Sect. 2.1.

It may seem surprising that the partition function of the multiplicative coalescent is so similar to that of the additive coalescent: near the start of the process, when most blocks have size 1, the additive coalescent has twice as many choices as the multiplicative coalescent. Later in the process, blocks should be larger, and one would guess that usually $xy > x + y$. Why these two effects should almost exactly cancel each other out is something of a mystery. On the other hand, the similarity of the partition functions may suggest that the additive and multiplicative coalescents have similar behaviour.

A more detailed investigation will reveal interesting behaviour whose subtleties are not captured by the above formulae. We will see in Sect. 2.3 that there is a naturally defined "empirical partition function" $\hat{Z}_{\mathrm{MC}}(n)$ such that $Z_{\mathrm{MC}}(n) = \mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n)\right]$. However, $\hat{Z}_{\mathrm{MC}}(n)$ is typically *exponentially smaller* than $Z_{\mathrm{MC}}(n)$ (see Corollary 4.3), so in a quantifiable sense, the partition function $Z_{\mathrm{MC}}(n)$ takes the value it does due to extremely rare events. Correspondingly, it turns out that the behaviour of the additive and multiplicative coalescents are typically quite different.

To analyze the typical value of $\hat{Z}_{\mathrm{MC}}(n)$, we are led to develop the connection between the multiplicative coalescent and the classical Erdős-Rényi random graph process $(G(n,p), 0 \leq p \leq 1)$. The most technical part of the notes is the proof of a concentration result for the susceptibility of $G(n,p)$; this is Theorem 4.4, below. Using a well-known coupling between the multiplicative coalescent and Kruskal's algorithm for the minimum weight spanning tree problem, our susceptibility bound leads easily to a novel proof of the $\zeta(3)$ limit for the total weight of the minimum spanning tree of the complete graph (this is stated in Theorem 5.1, below).[1]

1.1 Stylistic Remarks

The primary purpose of these notes is expository (though there are some new results, notably Theorems 4.2 and 4.4). Accordingly, we have often opted for repetition over concision. We have also included plenty of exercises and open problems (the open problems are mostly listed in Sect. 7). Some exercises state facts which are required later in the text; these are distinguished by a ⊛.

2 A Tale of Three Coalescents

2.1 Cayley's Formula and Pitman's Coalescent

We begin by describing the beautiful proof of Cayley's formula found by Jim Pitman, and its link with uniform spanning trees. Cayley's formula states that the number of trees with vertices $\{1, 2, \ldots, n\}$ is n^{n-2}, or equivalently that the number of *rooted* trees with vertices labeled by $[n] := \{1, 2, \ldots, n\}$ is n^{n-1}. To prove this formula, Pitman [13] analyzes a process we call Pitman's coalescent. To explain the process, we need some basic definitions. A *forest* is a graph with no cycles; its connected components are its trees. A *rooted forest* is a forest in which each tree t has a distinguished root vertex $r(t)$.

Pitman's Coalescent, Version 1. The process has n steps, and at step i consists of a rooted forest $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$ with $n+1-i$ trees. (At step

(continued)

---

[1] We find this proof of the $\zeta(3)$ limit for the MST weight pleasing, as it avoids lemmas which involve estimating the number of unicyclic and complex components in $G(n,p)$; morally, the cycle structure of components of $G(n,p)$ should be unimportant, since cycles are never created in Kruskal's algorithm!

1, these trees are simply isolated vertices with labels $1, \ldots, n$.) To obtain F_{i+1} from F_i, choose a pair (U_i, V_i), where $U_i \in [n]$ and V_i is the root of some tree of F_i not containing U_i, uniformly at random from among all such pairs. Add an edge from U_i to V_i, and root the resulting tree at the root of U_i's old tree. The forest F_{i+1} consists of this new tree together with the $n - i - 1$ unaltered trees from F_i.

The coalescents we consider all have the general form of Pitman's coalescent: they are forest-valued stochastic processes $(F_i, 1 \le i \le n)$, where $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$ is a forest with vertices labeled by $[n]$.

Pitman's Coalescent, Version 2. Consider the directed graph $K_n^{\rightarrow}$ with vertices $\{1, \ldots, n\}$ and an oriented edge from k to ℓ for each $1 \le k \ne \ell \le n$. Let $\mathbf{W} = \{W_{(k,\ell)} : 1 \le k \ne \ell \le n\}$ be independent copies of a continuous random variable W, that weight the edges of $K_n^{\rightarrow}$. Let F_1 be as in Version 1. For $i \in \{1, \ldots, n-1\}$, form F_{i+1} from F_i by adding the smallest weight edge (k, ℓ) whose head ℓ is the root of one of the trees in F_i. (Each tree of F_i is rooted at its unique vertex having indegree zero in F_i.)

Note that in Version 2, for each $i \in \{1, \ldots, n\}$ and each tree T of F_i, all edges of T are oriented away from a single vertex of T; so, viewing this vertex as the root of T, the orientation of edges in T is fully specified by the location of its root.

Exercise 1 View the trees of Version 2 as rooted rather than oriented. Then the sequences of forests $(F_1, \ldots, F_n)$ described in Version 1 and Version 2 have the same distribution.

Say that a finite set $\{X_i, i \in I\}$ of random variables is *exchangeable* if for any two deterministic orderings of I as, say, $i_1, \ldots, i_k$ and $i'_1, \ldots, i'_i$, the vectors $(X_{i_1}, \ldots, X_{i_k})$ and $(X_{i'_1}, \ldots, X_{i'_k})$ are identically distributed. In particular, if the elements of $\{X_i, i \in I\}$ are iid then the set is exchangeable.

Exercise 2 Suppose that the edge weights $\mathbf{W}$ are only assumed to be exchangeable and a.s. pairwise distinct. Show that the sequences of forests $(F_1, \ldots, F_n)$ described in Version 1 and Version 2 still have the same distribution.

To prove Cayley's formula, we compute the *partition function* of Pitman's coalescent: this is the total number of possibilities for its execution. (To do so, it's easiest to think about Version 1 of the procedure.) For example, when $n = 3$, there are 6 possibilities for the first step of the process: 3 choices for the first vertex, then 2 choices of a tree not containing the first vertex. For the second step, there are 3 choices for the first vertex; there is only one component not containing the chosen vertex, and we must choose it. Thus, for $n = 3$, the partition function has value $Z_{\mathrm{AC}}(3) = 6 \cdot 3 = 18$. More generally, for the n-vertex process, when adding the i'th edge we have n choices for the first vertex and $n - i$ choices of tree not containing the first vertex, so a total of $n(n - i)$ possibilities. Thus the partition function is

$$Z_{\mathrm{AC}}(n) = \prod_{i=1}^{n-1} n \cdot (n - i) = n^{n-1}(n - 1)! \qquad (2.1)$$

It is not possible to recover the entire execution path of the additive coalescent from the final tree, since there is no way to tell in which order the edges were added. If we wish to retain this information, we may label each edge of $T_1^{(n)}$ with the step at which it was added. More precisely, $L(e)$ is the unique integer $i \in \{1, \ldots, n - 1\}$ such that e is not an edge of F_i but is an edge of F_{i+1}. It follows from the definition of the process that the edge labels are distinct, so $L : E(T_1^{(n)}) \to \{1, \ldots, n - 1\}$ is a bijective map.

Now fix a rooted tree t with vertices $\{1, \ldots, n\}$, and consider the *restricted* partition function $Z_{\mathrm{AC},t}(n)$; this is simply the number of possibilities for the execution of the process for which the end result is the tree t. We claim that $Z_{\mathrm{AC},t}(n) = (n - 1)!$. This is easy to see: for any labelling ℓ of the edges of t with integers $\{1, \ldots, n - 1\}$, there is a unique execution path for which $(T_1^{(n)}, L) = (t, \ell)$, and there are $(n - 1)!$ possible labellings ℓ. Thus, the probability of ending with the tree t is $Z_{\mathrm{AC},t}(n)/Z_{\mathrm{AC}}(n) = 1/n^{n-1}$. Since this number doesn't depend on t, only on n, it follows that every rooted labelled tree with n vertices is equally likely, and so there must be n^{n-1} such trees.

Note *The preceding argument is correct, but treads lightly around an important point. When performing the process, the number of possibilities for the i'th edge does not depend on the first* $i - 1$ *choices, so the probability of building a particular tree t by adding its edges in a particular order is* $[n^{n-1}(n - 1)!]^{-1}$ *regardless of the order. Of course, the* set *of possible choices at a given step must depend on the history of the process – for example, we must not add a single edge twice. More generally, thinking of Version 2, applying the procedure to a graph other than* $K_n^{\to}$ *need not yield a uniform spanning tree of the graph, and indeed may not even build a tree. (Consider, for example, applying the procedure to a two-edge path.)*

By stopping Pitman's coalescent before the end, one can use a similar analysis to obtain counting formulae for forests. Write $Z_{\mathrm{AC}}(n, k)$ for the total number of possibilities for Pitman's coalescent stopped at step k (so ending with $n + 1 - k$ forests). We write $(m)_\ell$ to denote the *falling factorial* $\prod_{i=0}^{\ell-1}(m - i)$.

Exercise 3

(a) Show that $Z_{\mathrm{AC}}(n, k) = n^{k-1}(n - 1)_{k-1}$ for each for $1 \le k \le n$.
(b) An *ordered labeled forest* is a sequence $(t_1, \ldots, t_\ell)$ where each t_i is a rooted labelled tree and all labels of vertices in the forest are distinct. Show that

for each $1 \le \ell \le n$ the number of ordered labeled forests $(t_1, \dots, t_\ell)$ with $\bigcup_{i=1}^{\ell} V(t_i) = [n]$, is $\ell \cdot n^{n-\ell-1} \cdot (n)_\ell$.

We briefly discuss a special case of Version 2. Suppose that $W_{(k,l)}$ is exponential with rate $X_{(k,\ell)}$, where $\mathbf{X} = \{X_{(k,\ell)} : 1 \le k \ne \ell \le n\}$ are independent copies of any non-negative random variable X. By standard properties of exponentials and the symmetry of the process, the dynamics in this case may be described as follows.

Pitman's Coalescent, Version 3. Let F_1 be as in Version 1. For $i \in \{1, \dots, n-1\}$, choose an edge whose head is the root of any one of the trees in F_i, each such edge (k, l) chosen with probability proportional to its weight $X_{(k,l)}$; add the chosen edge to create the forest F_{i+1}.

Consider Version 3 of the procedure after $i - 1$ edges have been added. Conditional on $\mathbf{X}$ and on the forest $(T_1^{(i)}, \dots, T_{n-i+1}^{(i)})$, the probability of adding a particular edge (k, ℓ) whose head is a root, is proportional to $X_{(k,\ell)}$, so is equal to

$$\frac{X_{(k,\ell)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\dots,n\} \setminus V(T_m^{(i)})} X_{(j, r(T_m^{(i)}))}} .$$

Now fix any sequence $f_1, \dots, f_n$ of forests that can arise in the process. Write $f_i = (t_k^{(i)}, 1 \le k \le n + 1 - i)$ and for $i = 1, \dots, n-1$ write (k_i, ℓ_i) for the unique edge of f_{i+1} not in f_i. Then by the above,

$$\mathbf{P}\{F_i = f_i, 1 \le i \le n \mid \mathbf{X}\} = \prod_{i=1}^{n-1} \frac{X_{(k_i,\ell_i)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\dots,n\} \setminus V(t_m^{(i)})} X_{(j, r(t_m^{(i)}))}} .$$

By Exercise 1 and the above analysis, it follows that for any such sequence $f_1, \dots, f_n$,

$$\mathbf{E}\left[\prod_{i=1}^{n-1} \frac{X_{(k_i,\ell_i)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\dots,n\} \setminus V(t_m^{(i)})} X_{(j, r(t_m^{(i)}))}} \right] = \frac{1}{n^{n-1}(n-1)!} .$$

It is by no means obvious at first glance that this expectation should not depend on law of X, let alone that it should have such a simple form.

2.2 *Kingman's Coalescent and Random Recursive Trees*

Pitman's coalescent starts from isolated vertices labeled from $\{1, \dots, n\}$, and builds a rooted tree by successive edge addition. At each step, an edge is added *to* some vertex, *from* some root (of a component not containing the chosen vertex). When we calculated $Z_{\mathrm{AC}}(n)$, it was important that the number of possibilities at each step depended only on the *number* of trees in the current forest and not, say, their sizes, or some other feature.

Pitman's merging rule (*to* any vertex, *from* a root) yielded a beautiful proof of Cayley's formula. It is natural to ask what other rules exist, and what information may be gleaned from them. Of course, *from* any vertex, *to* a root just yields the additive coalescent, with edges of the resulting tree oriented towards the root rather than towards the leaves. What about *from* any root, *to* any (other) root, as in the following procedure? In a very slight abuse of terminology, we call this rule *Kingman's coalescent*. We again start from a rooted forest F_1 of n isolated vertices $\{1, \ldots, n\}$. Recall that we write $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$.

Kingman's Coalescent. At step i, choose an ordered pair (U_i, V_i) of distinct roots from $\{r(T_1^{(i)}), \ldots, r(T_{n+1-i}^{(i)})\}$, uniformly at random from among the $(n+1-i)(n-i)$ such pairs. Add an edge from U_i to V_i, and root the resulting tree at U_i. The forest F_{i+1} consists of this new tree together with the $n-i-1$ unaltered trees from F_i.

Our convention is that when an edge is added from u to v, the root of the resulting tree is u; this maintains that edges are always oriented towards the leaves. For Kingman's coalescent, when i trees remain there are $i(i-1)$ possibilities for which oriented edge to add. Like for Pitman's coalescent, this number depends only on the number of trees, and it follows that the total number of possible execution paths for the process is

$$Z_{\mathrm{KC}}(n) = \prod_{i=2}^{n} i(i-1) = n!(n-1)!\,. \tag{2.2}$$

What does this number count?

To answer the preceding question, as in the additive coalescent let $L : E(T_1^{(n)}) \to \{1, \ldots, n-1\}$ label the edges of $T_1^{(n)}$ in their order of addition. It is easily seen that for Kingman's coalescent, the edge labels decrease along any root-to-leaf path; we call such a labelling a *decreasing edge labelling*.[2] Furthermore, any decreasing edge labelling of $T_1^{(n)}$ can arise. Once again, the full behaviour of the coalescent is described by pair $(T_1^{(n)}, L)$, and conversely, the coalescent determines $T_1^{(n)}$ and L. These observations yield that the number of rooted trees with vertices labelled

---

[2] It is more common to order by *reverse* order of addition, so that labels increase along root-to-leaf paths; this change of perspective may help with Exercise 4.

$\{1,\ldots,n\}$, additionally equipped with a decreasing edge labelling, is $n!(n-1)!$. The factor $n!$ simply counts the number of ways to assign the labels $\{1,\ldots,n\}$ to the vertices. By symmetry, each vertex labelling of a given tree is equally likely to arise, and so we have the following.

Proposition 2.1 *The number of pairs* (T,L)*, where* T *is a rooted tree with* n *vertices and* L *is a decreasing edge labelling of* T*, is* $(n-1)!$.

Exercise 4 (Random recursive trees) Prove Proposition 2.1 by introducing and analyzing an n-step procedure that at step i constructs a rooted tree with i vertices.

Before the next exercise, we state a few definitions. For a graph G, write $|G|$ for the number of vertices of G. If T is a rooted tree and u is a vertex of T, write T_u for the subtree of T consisting of u together with its descendants in T (we call T_u the subtree of T rooted at u). Also, if u is not the root, write $p(u)$ for the parent of u in T. Finally, write $\mathrm{aut}(T)$ for the number of rooted automorphisms of T.

Exercise 5 Show that for a fixed rooted tree T, the number of decreasing edge labellings of T with labels $1,2,\ldots,|V(T)|-1$ is

$$\frac{1}{\mathrm{aut}(T)}\cdot\prod_{v\in V(T)}\frac{(|T_v|-1)!}{\prod_{\{u\in V(T):p(u)=v\}}|T_u|!}\,.$$

Our convention is that an empty product equals 1; a special case is that $0!=1$. It follows from the preceding exercise that, writing $\mathcal{T}_n$ for the set of rooted trees with n vertices,

$$\sum_{T\in\mathcal{T}_n}\frac{1}{\mathrm{aut}(T)}\cdot\prod_{v\in V(T)}\frac{|E(T_u)|!}{\prod_{\{u\in V(T):p(u)=v\}}|V(T_u)|!}=(n-1)!\,;$$

another formula that one may not find obvious at first glance.

To finish the section, note that just like for Pitman's coalescent, we might well consider a version of this procedure that is "driven by" iid non-negative weights $\mathbf{X}=\{X_{(k,\ell)}:1\le k\ne\ell\le n\}$. (Recall that we viewed these weights as exponential *rates*, then used the resulting exponential clocks at each step to determine which edge to add.) At each step, add an oriented edge whose tail and head are both the roots of some tree of the current forest, each such edge chosen with probability proportional to its weight. For this procedure, conditional on $\mathbf{X}$, after adding the first $i-1$ edges, the conditional probability of adding a particular edge (k,ℓ) is

$$\frac{X_{(k,\ell)}}{\sum_{1\le j\ne m\le n}X_{(r(T_j^{(i)}),r(T_m^{(i)}))}}\,.$$

Now fix any sequence $f_1, \ldots, f_n$ of forests that can arise in the process, write $f_i = (t_k^{(i)}, 1 \le k \le n+1-i)$, and for $i = 1, \ldots, n-1$ write (k_i, ℓ_i) for the unique edge of f_{i+1} not in f_i. Then we have

$$\mathbf{P}\{F_i = f_i, 1 \le i \le n \mid \mathbf{X}\} = \prod_{i=1}^{n-1} \frac{X_{(k_i,\ell_i)}}{\sum_{1 \le m \ne j \le n} X_{(r(t_m^{(i)}), r(t_j^{(i)}))}} .$$

It follows from the above analysis that for any such sequence $f_1, \ldots, f_n$,

$$\mathbf{E}\left[\prod_{i=1}^{n-1} \frac{X_{(k_i,\ell_i)}}{\sum_{1 \le m \ne j \le n} X_{(r(t_m^{(i)}), r(t_j^{(i)}))}}\right] = \frac{1}{n!(n-1)!} .$$

Once again, it is not even a priori clear that this expectation should not depend on the law of X.

Exercise 6 (First-passage percolation) Develop and analyze a "Version 3" variant of the tree growth procedure from Exercise 4, using exponential edge weights.

2.3 *The Multiplicative Coalescent and Minimum Spanning Trees*

The previous two sections considered merging rules of the form any-to-root and root-to-root, and obtained Pitman's coalescent and Kingman's coalescent, respectively. We now take up the "any-to-any" merging rule. This is arguably the most basic of the three rules, but its behaviour is arguably the hardest to analyze. We begin as usual from a forest F_1 of n isolated vertices $\{1, \ldots, n\}$, and write $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$. In the multiplicative coalescent there is no natural way to maintain the property that edges are oriented toward some root vertex, so we view the trees of the forests as unrooted, and their edges as unoriented. Given a set S, write $\binom{S}{k}$ for the set of k-element subsets of S.

The multiplicative coalescent. To obtain F_{i+1} from F_i, choose an pair $\{U_i, V_i\}$ uniformly at random from the set of pairs $\{u, v\} \in \binom{[n]}{2}$ for which u and v are different trees of F_i. Add an edge from U_i to V_i to form the forest F_{i+1}.

This is known as the *multiplicative* coalescent, because the number of possible choices of an edge joining trees $T_j^{(i)}$ and $T_k^{(i)}$ is $|T_j^{(i)}||T_k^{(i)}|$. It follows that the number of possible edges that may be added to the forest F_i is

$$\sum_{1\le j\neq k\le n+1-i} |T_j^{(i)}||T_k^{(i)}| = \frac{1}{2}\left(n^2 - \sum_{T\in F_i} |T|^2\right).$$

The above expression is more complicated than for the additive coalescent or Kingman's coalescent: it depends on the forest F_i, for one.

In much of the remainder of these notes, we investigate an expression for the partition function $Z_{\mathrm{MC}}(n)$ of the multiplicative coalescent that arises from the preceding formula. To obtain this expression, recall the definition of an n-chain from Sect. 1, and that $\mathcal{P}_n$ is the set of n-chains.

Exercise 7 Show that $|\mathcal{P}_n| = \frac{(n!)^2}{n\cdot 2^{n-1}}$.

The multiplicative coalescent determines an n-chain in which the i'th partition is simply $P(F_i) := \{V(T_j^{(i)}), 1 \le j \le n+1-i\}$. It is straightforward to see that the number of possibilities for the multiplicative coalescent that give rise to a particular n-chain $P = (P_1, \ldots, P_n)$ is simply

$$\prod_{i=1}^{n-1} |A_i(P)||B_i(P)|,$$

where $A_i(P)$ and $B_i(P)$ are the parts of P_i that are combined in P_{i+1}. It follows that

$$Z_{\mathrm{MC}}(n) = \sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \prod_{i=1}^{n-1} \left(|A_i(P)||B_i(P)|\right).$$

This certainly looks more complicated than in the previous two cases. However, there is an exact formula for $Z_{\mathrm{MC}}(n)$ whose derivation is perhaps easier than for either $Z_{\mathrm{AC}}(n)$ or $Z_{\mathrm{KC}}(n)$ (though it does rely on Cayley's formula).

Proposition 2.2 $Z_{\mathrm{MC}}(n) = n^{n-2}(n-1)!$

Proof Let $\mathcal{S}$ be the set of pairs (t, ℓ) where t is an unrooted tree with $V(t) = [n]$ and $\ell : E(t) \to [n-1]$ is a bijection. By Cayley's formula, the number of trees t with $V(t) = [n]$ is n^{n-2}, so $\mathcal{S} = n^{n-2}(n-1)!$.

For $e \in E(T_n^{(1)})$, let $L(e) = \sup\{i : e \notin E(F_i)\}$. Then $L : E(T_n^{(1)}) \to [n-1]$ is a bijection. Thus the pair $(T_n^{(1)}, L)$ is an element of $\mathcal{S}$. To see this map is bijective, note that if $(T_n^{(1)}, L) = (t, \ell)$ then for each $1 \le i \le n$, F_i is the forest on $[n]$ with edges $\{\ell^{-1}(j), 1 \le j < i\}$. The result follows. □

The above proposition yields that $Z_{\mathrm{MC}}(n) = Z_{\mathrm{AC}}(n)/n$. If we were to additionally choose a root for $T_n^{(1)}$, we would obtain identical partition functions. This suggests that perhaps the additive and multiplicative coalescents have similar structures. One might even be tempted to believe that the trees built by the two coalescents are identically distributed; the following exercise (an observation of Aldous [3]), will disabuse you of that notion.

Exercise 8 Let T be built by the multiplicative coalescent, and let T' be obtained from the additive coalescent by unrooting the final tree. Show that if $n \geq 4$ then T and T' are not identically distributed.

Despite the preceding exercise, it is tempting to guess that the two trees are still similar in structure; this was conjectured by Aldous [3], and only recently disproved [2]. In the remainder of the section, we begin to argue for the difference between the two coalescents, from the perspective of their partition functions. For $1 \leq k \leq n$, write $Z_{\mathrm{MC}}(n,k)$ for the partition function of the first k steps of the multiplicative coalescent,

$$Z_{\mathrm{MC}}(n,k) = \sum_{P=(P_1,\ldots,P_k)\in\mathcal{P}_{n,k}} \prod_{i=1}^{k-1} \left(|A_i(P)||B_i(P)|\right),$$

where $\mathcal{P}_{n,k}$ is the set of length-k initial segments of n-chains. We have, e.g., $Z_{\mathrm{MC}}(n,1) = 1$, $Z_{\mathrm{MC}}(n,2) = \binom{n}{2}$, and $Z_{\mathrm{MC}}(n,n) = Z_{\mathrm{MC}}(n)$.

The argument of Proposition 2.2 shows that $Z_{\mathrm{MC}}(n,k) = u_{n,k} \cdot (k-1)!$, where $u_{n,k}$ is the number of unrooted forests with vertices $[n]$ and $k-1$ total edges. The identity

$$u_{n,k} = \binom{n}{n+1-k} n^{k-2} \sum_{i=0}^{n+1-k} \left(\frac{-1}{2n}\right)^i \binom{n+1-k}{i} (n+1-k+i)\cdot(k-1)_i,$$

was derived by Rényi [14], and I do not know of an exact formula that simplifies the above expression. We begin to see that there is more to the multiplicative coalescent than first meets the eye.

If we can't have a nice, simple identity, what about bounds? Of course, there is the trivial upper bound $Z_{\mathrm{MC}}(n,k) \leq (n(n-1)/2)^{k-1}$, since at each step there are at most $\binom{n}{2}$ pairs to choose from; similar bounds hold for the other two coalescents. To improve this bound, and more generally to develop a deeper understanding of the dynamics of the multiplicative coalescent, our starting point is the following observation.

Given an n-chain $P = (P_1, \ldots, P_n)$, for the multiplicative coalescent we have

$$\mathbf{P}\left\{(P(F_i), 1 \leq i \leq n) = P\right\} = \prod_{i=1}^{n-1} \frac{2|A_i(P)||B_i(P)|}{n^2 - \sum_{\pi\in P_i} |\pi|^2}.$$

This holds since for $1 \le i \le n-1$, given that $P(F_j) = P_j$ for $1 \le j \le i$, there are $(n^2 - \sum_{\pi \in P_i} |\pi|^2)/2$ choices for which oriented edge to add to form F_{i+1}, and $P(F_{i+1}) = P_{i+1}$ for precisely $|A_i(P)||B_i(P)|$ of these. It follows that

$$\begin{aligned} Z_{\mathrm{MC}}(n) &= \sum_{P=(P_1,\ldots,P_n)\in \mathcal{P}_n} \mathbf{P}\{(P(F_i), 1 \le i \le n) = P\} \cdot \prod_{i=1}^{n-1} \frac{n^2 - \sum_{\pi \in P_i} |\pi|^2}{2} \\ &= \sum_{P=(P_1,\ldots,P_n)\in \mathcal{P}_n} \mathbf{P}\{(P(F_i), 1 \le i \le n) = P\} \cdot 2^{-(n-1)} \\ &\qquad \cdot \mathbf{E}\left\{ \prod_{i=1}^{n-1} \left(n^2 - \sum_{T \in F_i} |T|^2 \right) \,\middle|\, (P(F_i), 1 \le i \le n) = P \right\} \\ &= 2^{-(n-1)} \cdot \mathbf{E}\left[\prod_{i=1}^{n-1} \left(n^2 - \sum_{T \in F_i} |T|^2 \right) \right]. \end{aligned} \tag{2.3}$$

A mechanical modification of the logic leading to (2.3) yields the following expression, valid for each $1 \le k \le n$:

$$Z_{\mathrm{MC}}(n,k) = 2^{-(k-1)} \mathbf{E}\left[\prod_{i=1}^{k-1} \left(n^2 - \sum_{T \in F_i} |T|^2 \right) \right]. \tag{2.4}$$

Write

$$\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k) = \prod_{i=1}^{k-1} \left(n^2 - \sum_{T \in F_i} |T|^2 \right),$$

let $\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n) = \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,1)$, and let $\hat{Z}_{\mathrm{MC}}(n,k) = 2^{-(k-1)} \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k)$ and $\hat{Z}_{\mathrm{MC}}(n) = \hat{Z}_{\mathrm{MC}}(n,n)$. With this notation, (2.3) and the subsequent equation state that

$$\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n,k) \right] = Z_{\mathrm{MC}}(n,k) = \frac{1}{2^{k-1}} \mathbf{E}\left[\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k) \right]. \tag{2.5}$$

The random variable $\hat{Z}_{\mathrm{MC}}(n)$ is a sort of *empirical partition function* of the multiplicative coalescent. The superscript arrow on $\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k)$ is because the factor 2^{k-1} may be viewed as corresponding to a choice of orientation for each edge of F_k. The random variable $\hat{Z}_{\mathrm{MC}}(n)$ of course contains more information than simply its expected value, so by studying it we might hope to gain a greater insight into the behaviour of the coalescent. Much of the remainder of these notes is devoted to showing that $\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n) \right] = Z_{\mathrm{MC}}(n)$ is a *terrible* predictor of the typical value of $\hat{Z}_{\mathrm{MC}}(n)$. More precisely, there are unlikely execution paths along which the multiplicative coalescent has many more possibilities than along a typical path; such paths swell the *expected* value of $\hat{Z}_{\mathrm{MC}}(n)$ to exponentially larger than its *typical* size.

The logic leading to (2.3) and (2.4) may also be applied to the additive coalescent; the result is boring but instructive. First note that

$$Z_{\mathrm{AC}}(n,k) = \sum_{P=(P_1,\ldots,P_k)\in\mathcal{P}_{n,k}} \prod_{i=1}^{k-1} (|A_i(P)| + |B_i(P)|)\,.$$

For the additive coalescent, the total number of choices at step i is $n(n-i)$, and given that $P(F_i) = P_i$, the number of choices which yield $P(F_{i+1}) = P_{i+1}$ is $A_i(P)+B_i(P)$. Writing $\mathbf{P}_{\mathrm{AC}}$ for probabilities under the additive coalescent, we thus have

$$\mathbf{P}_{\mathrm{AC}}\left\{(P(F_i), 1 \le i \le k) = (P_1,\ldots,P_k)\right\} = \prod_{i=1}^{k-1} \frac{|A_i(P)| + |B_i(P)|}{n(n-i)}$$

Following the logic through yields

$$Z_{\mathrm{AC}}(n,k) = \mathbf{E}_{\mathrm{AC}}\left[\prod_{i=1}^{k-1} n(n-i)\right] = \mathbf{E}_{\mathrm{AC}}\left[n^{k-1}(n-1)_{k-1}\right].$$

Thus, the "empirical partition function" of the additive coalescent is a constant, so contains no information beyond its expected value. (This fact is essentially the key to Pitman's proof of Cayley's formula.)

The terms of the products (2.3) and (2.4), though random, turn out to behave in a very regular manner (but proving this will take some work). Through a study of these terms, we will obtain control of $\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n)\right]$, and thereby justify the above assertion that $\hat{Z}_{\mathrm{MC}}(n)$ is typically very different from its mean.

2.3.1 The Growth Rate of $Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)$

As a warmup, and to introduce a key tool, we approximate the value of $Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)$ using a connection between the multiplicative coalescent and a process we call (once again with a very slight abuse of terminology) the Erdős-Rényi coalescent. Write K_n for the *complete graph*, i.e. the graph with vertices $[n]$ and edges $(\{i,j\}, 1 \le i < j \le n)$.

The Erdős-Rényi coalescent. Choose a uniformly random permutation $e_1,\ldots,e_{\binom{n}{2}}$ of $E(K_n)$. For $0 \le i \le \binom{n}{2}$, let $G_i^{(n)}$ have vertices $[n]$ and edges $\{e_1,\ldots,e_i\}$.

Our indexing here starts at zero, unlike in the multiplicative coalescent; this is slightly unfortunate, but it is standard for the Erdős-Rényi graph process to index so that $G_i^{(n)}$ has i edges. This process is different from the previous coalescent processes, most notably because it creates graphs with cycles.

Note that we can recover the multiplicative coalescent from the Erdős-Rényi coalescent in the following way. Informally, simply ignore any edges added by the Erdős-Rényi coalescent that fail to join distinct components. More precisely, for each $0 \leq m \leq \binom{n}{2}$, let τ_m be the number of edges $\{U_i, V_i\}$, $0 < i \leq m$ such that U_i and V_i lie in different components of $G_{i-1}^{(n)}$. (See Fig. 1 for an example.)

Observe that

$$\tau_m + 1 = \begin{cases} \tau_m & \text{if } G_{m+1}^{(n)} \text{ and } G_m^{(n)} \text{ have the same number of components} \\ \tau_m + 1 & \text{if } G_{m+1}^{(n)} \text{ has one fewer component than } G_m^{(n)} \,. \end{cases}$$

In other words, τ_m increases precisely when the endpoints of the edge added to $G_m^{(n)}$ are in different components. Further, the set

$$\{e_m : m \geq 1, \tau_m > \tau_{m-1}\}$$

contains $n-1$ edges, since $G_0^{(n)}$ has n components and $G_{\binom{n}{2}}^{(n)}$ almost surely has only one component.

Set $I_1 = 0$ and for $1 < k \leq n$ let

$$I_k = \inf\{m \geq 1 : \tau_m = k-1\} \,.$$

Then for $1 < k \leq n$, the edge e_{I_k} joins distinct components of $G_{I_k-1}^{(n)}$, and by symmetry is equally likely to be any such edge. Thus, letting F_k be the graph with edges $\{e_{I_j} : 1 \leq j \leq k\}$ for $1 \leq k \leq n$, the process $\{F_k, 1 \leq k \leq n\}$ is precisely distributed as the multiplicative coalescent. This is a *coupling* between the Erdős-Rényi graph process and the multiplicative coalescent; its key property is that for all

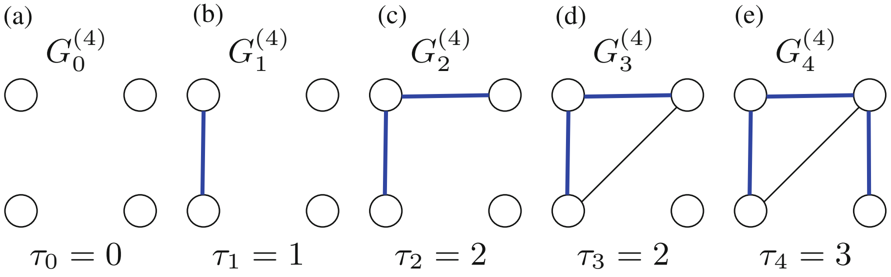

Fig. 1 An example of the first steps of the Erdős-Rényi coalescent. The multiplicative coalescent is obtained by keeping only the *thicker, blue edges*

$1 \le k \le n$, the vertex sets of the trees of F_k are the same as those of the components of $G^{(n)}_{I_k}$.

Having found the multiplicative coalescent within the Erdős-Rényi coalescent, we can now use known results about the latter process to study the former. For a graph G, and $v \in V(G)$, we write $N(v) = N_G(v)$ for the set of nodes adjacent to v (the *neighbours* of v), and write $C(v) = C_G(v)$ for the connected component of G containing v. We will use the results of the following exercise.[3]

Exercise 9 ⊛

(a) Show that in the Erdős-Rényi coalescent, if all components have size at most s then the probability a uniformly random edge from among the remaining edges has both endpoints in the same component is at most $(s-1)/(n-1)$.
(b) Show that for all $0 \le m \le n/2$, in $G^{(n)}_m$, $\mathbf{E}|N(v)| \le 2m/n$.
(c) Prove by induction that for all $0 \le m < n/2$, in $G^{(n)}_m$, $\mathbf{E}\left[|C(1)|\right] \le n/(n-2m)$.
(**Hint.** First condition on $N(1)$, then average.)
(d) Prove that for all $\epsilon > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(G^{(n)}_{(1-\epsilon)n/2} \text{ has a component of size } > \epsilon n\right) = 0\,.$$

(**Hint.** Given that the largest component of $G^{(n)}_m$ has size s, with probability at least s/n vertex 1 is in such a component.)

Using the above exercise, we now fairly easily prove a lower bound on the partition function of the first half of the multiplicative coalescent.

Proposition 2.3 *For all $\beta > 0$,*

$$\mathbf{P}\left\{\hat{Z}^{\rightarrow}_{\mathrm{MC}}(n, \lfloor n/2 \rfloor) \ge n^{(1-\beta)n}\right\} \to 1 \ \textit{as}\ n \to \infty\,.$$

We begin by showing that typically $I_t = (1+o(1))t$ until $t \ge n/2$.

Lemma 2.4 *For all $\epsilon > 0$, $\limsup_{n\to\infty} \mathbb{P}\left(I_{(1-\epsilon)n/2} \ge n/2\right) = 0$.*

Proof Fix $\epsilon > 0$, let $\delta = \epsilon/3$, and let E be the event that all components of $G_{(1-\delta)n/2}$ have size at most δn. For $m \ge 0$, conditional on $G^{(n)}_m$, by Exercise 9(a), $\tau_{m+1} - \tau_m$ stochastically dominates a Bernoulli$(1-(s-1)/(n-1))$ random variable, where s is the size of the largest component of $G^{(n)}_m$.

For n large and $s \le \delta n$ we have $1-(s-1)/(n-1) \ge 1-\epsilon/2$. Therefore, on E and for large n the sequence $(\tau_{m+1}-\tau_m, 0 \le m < (1-\delta)n/2)$ stochastically dominates a sequence $(B_m, 0 \le m < (1-\delta)n/2)$ of iid Bernoulli$(1-\epsilon/2)$ random

---

[3] Until further notice, we omit ceilings and floors for readability.

variables. It follows that

$$\begin{aligned}\mathbf{P}\left\{\tau_{(1-\delta)n/2} \le (1-\epsilon)n/2\right\} &\le \mathbf{P}\left\{E^c\right\} + \mathbf{P}\left\{\tau_{(1-\delta)n/2} \le (1-\epsilon)n/2 \mid E\right\} \\ &\le \mathbf{P}\left\{E^c\right\} + \mathbf{P}\left\{\mathrm{Bin}((1-\delta)n/2, 1-\epsilon/2) < (1-\epsilon)n/2\right\} \\ &= o(1)\,,\end{aligned}$$

the last line Exercise 9(d) and Chebyshev's inequality (note that $(1-\delta)(1-\epsilon/2)n/2 > (1-5\epsilon/6)n/2$). On the other hand, if $\tau_{(1-\delta)n/2} > (1-\epsilon)n/2$ then $I_{(1-\epsilon)n/2} \le (1-\delta)n/2 < n/2$. ☐

Proof of Proposition 2.3 View $(F_1, \ldots, F_n)$ as coupled with the Erdős-Rényi coalescent as above, so that F_k and $G^{(n)}_{I_k}$ have the same components. Fix $\delta \in (0, 1/4)$ and let $k = k(n) = n/2 - 2\delta n$. Let E_1 be the event that $I_{n/2-\delta n} < n/2$.[4] Since $I_{m+1} \ge I_m + 1$ for all m, we have

$$I_k \le I_{n/2-\delta n} - ((n/2 - \delta n) - k) = I_{n/2-\delta n} - \delta n\,.$$

Thus, on E_1 we have $I_k \le (1-2\delta)n/2$.

Next let E_2 be the event that all component sizes in $G^{(n)}_{(1-2\delta)n/2}$ are at most δn. The components of F_k are precisely the components of $G^{(n)}_{I_k}$, so if $E_1 \cap E_2$ occurs then since on E_1 we have $I_k \le (1-2\delta)n/2$, all components of F_k have size at most δn. In this case, for all $i \le k$ the components of F_i clearly also have size at most δn.

It follows[5] that on $E_1 \cap E_2$, for all $i \le k$,

$$\sum_{T \in F_i} |T|^2 \le \delta n^2$$

so on $E_1 \cap E_2$,

$$\begin{aligned}\hat{Z}_{\overrightarrow{\mathrm{MC}}}(n, k+1) &= \prod_{i=1}^{k}\left(n^2 - \sum_{T \in F_i} |T|^2\right) \\ &\ge n^{2k}(1-\delta)^k \qquad\qquad (2.6) \\ &= n^{n(1-4\delta)} \cdot (1-\delta)^{n/2} \qquad > n^{n(1-5\delta)}\,,\end{aligned}$$

the last inequality holding for n large. By Exercise 9(d) and Lemma 2.4, $\mathbb{P}(E_1 \cap E_2) \to 1$ as $n \to \infty$. Since $\hat{Z}_{\overrightarrow{\mathrm{MC}}}(n, \lfloor n/2 \rfloor) \ge \hat{Z}_{\overrightarrow{\mathrm{MC}}}(n, k+1)$ for n large, the result follows. ☐

The following exercise is to test whether you are awake.

---

[4] We omit the dependence on n in the notation for E_1; similar infractions occur later in the proof.

[5] To maximize $\sum_j x_j^2$ subject to the conditions that $\sum_j x_j = 1$ and that $\max_j x_j \le \delta$, take $x_j = \delta$ for $1 \le j \le \delta^{-1}$.