

Distribution de valuations sur les arbres.

Michel Nguyên-Thê

▶ **To cite this version:**

Michel Nguyên-Thê. Distribution de valuations sur les arbres.. Informatique [cs]. Ecole Polytechnique X, 2004. Français. NNT : . pastel-00000839

HAL Id: pastel-00000839
https://pastel.hal.science/pastel-00000839

Submitted on 21 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée pour obtenir le grade de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité :

INFORMATIQUE

par

Michel NGUYEN THE

Titre de la thèse :

DISTRIBUTIONS DE VALUATIONS SUR LES ARBRES

Soutenue le lundi 9 février 2004 devant le jury composé de :

M.	Philippe FLAJOLET	Président
Mme	Danièle GARDY	Rapporteurs
M.	Guy LOUCHARD	
Mme	Mireille BOUSQUET-MÉLOU	Examinateurs
M.	Conrado MARTÍNEZ	
M.	Jean-Marc STEYAERT	(Directeur)

"For an academic career puts a young man into a kind of embarrassing position by requiring him to produce scientific publications in impressive quantity — a seduction into superficiality which only strong characters are able to withstand. Most practical occupations, however, are of such a nature that a man of normal ability is able to accomplish what is expected of him. His day-to-day existence does not depend on any special illuminations. If he has deeper scientific interests he may plunge into his favorite problems in addition to doing his required work. He needs not be oppressed by the fear that his efforts may lead to no results."

Albert Einstein

Remerciements

À Charles Alunni pour avoir créé un séminaire Pensée des sciences et pour m'avoir offert un exemplaire d'un numéro de la Revue de Synthèse.

À François Anceau que j'ai croisé non seulement à l'X au Magnan mais aussi à divers séminaires extra-académiques.

À Pierre Arnoux, mon ancien maître de stage en licence, que j'ai croisé plusieurs fois pendant ma thèse, à l'X, à Paris ou à Marseille. Grâce à lui j'ai un exemplaire papier du poly de l'X sur les systèmes dynamiques.

À Amine A. pour sa compétence et le repas qu'il m'a offert dans un restaurant libanais.

À Cyril Banderier pour de multiples services. Grâce à lui j'ai eu connaissance du *Sixth Seminar on Analysis of Algorithms* et j'ai pu m'y inscrire à temps à un jour près. Il m'a aussi entre autres donné un exemplaire de la thèse de Michèle Soria retapée en Latex à l'occasion. Je voudrais aussi le féliciter pour son travail autour du séminaire *Discrete Random Walks 2003* à l'Institut Henri Poincaré.

À Fabrice Baronnet pour sa compétence et son écoute.

À Frédérique Bassino pour m'avoir fait prendre en ATER à l'université de Marne-la-Vallée.

À Marie-Pierre Béal pour avoir assisté à mon séminaire à l'IGM un jour de vaches maigres (voir Crochemore).

À Behshad Behzadi pour m'avoir fait partager ses connaissances en algorithmique.

À Adrian Belu pour ses conversations de haute voltige.

À Catherine Bensoussan pour son efficacité et pour son humeur toujours posée et conciliante.

À Jean Berstel pour une très intéressante discussion au CIRM, ainsi que pour m'avoir donné un exemplaire de son article sur les séries formelles et parlant notamment du jeu d'opérateurs $+$ et $\times$, et pour avoir assisté à ma présentation au séminaire de l'IGM.

À Nicolas Billaudé pour son intérêt pour l'épistémologie des sciences.

À Florence Bister pour m'avoir aidé à préparer mes cours à Marne-la-Vallée.

À Jérémie Bourdon pour la logistique autour de la connexion Internet à Tatihou lors du *Seventh Seminar on Analysis of Algorithms.*

À mon co-bureau ~~Dimitri~~ Simon Bliudze pour d'innombrables discussions sur des sujets divers et variés.

À Mireille Bousquet-Mélou pour son attention pendant les séminaires.

À Pierre Cartier pour ses discussions toujours très intéressantes et très profondes pendant et après ses séminaires.

À celui qui lit la ligne en-dessous.

À celui qui lit cette ligne.

À celui qui lit la ligne en-dessus.

À Philippe Chassignet pour m'avoir recruté dans l'équipe des superviseurs du South Western Europe Regional ACM Programming Constest 2003.

À Brigitte Chauvin pour m'avoir permis d'assister à son cours de DEA en commun avec Danièle Gardy, et pour être venue assister à ma soutenance.

À Frédéric Chyzak pour sa relecture si attentive de mes résumés de séminaire Algorithmique de l'INRIA, et aussi pour m'avoir guidé dans l'utilisation de Maple et des packages INRIA.

Au CIRM pour leur accueil de très haute qualité pour les colloques Aléa.

À Julien Clément pour m'avoir donné l'idée de travailler sur la compactification des tries et pour avoir transféré son ordinateur fixe vers mon bureau à Marne-la-Vallée.

À Virginie Collette pour m'avoir donné un badge INRIA et me donner l'autorisation d'avoir un compte informatique au laboratoire Algo.

À Alain Connes pour m'avoir permis d'assiter à la cérémonie des Clay mathematics.

À Robert Cori pour m'avoir redirigé vers Jean-Marc Steyaert à l'issue de mon DEA.

À Radhia Cousot pour diverses conversations et pour les conseils qu'elle m'a donnés.

À Maxime Crochemore pour son chaleureux accueil à l'Institut Gaspard Monge lors de mon arrivée en ATER. Il a toujours fait montre d'une grande attention et d'une grande diplomatie, m'expliquant notamment les difficultés d'acheminement de nouveaux ordinateurs dont l'un m'était destiné, et aussi présentant ses excuses pour l'absentéisme généralisé lors de ma ma présentation au séminaire de l'IGM, dû à une réunion concomitante.

À Arnaud Dartois pour l'organisation du séminaire des thésards du LIX.

À Vincent David pour avoir repoussé les envahisseurs pendant que je faisais tranquillement ma thèse et pour son travail autour du développement durable.

À Pierre Demers pour ses travaux et pour les repas partagés ensemble sur Paris.

À Alain Denise pour m'avoir donné un exemplaire de sa thèse.

À Luc Devroye pour ses images d'arbres que je n'ai malheureusement pas pu exploiter à fond pour ma soutenance de thèse, faute de pouvoir afficher du postscript suffisamment rapidement.

À Michael Drmota pour son professionnalisme dans l'organisation du Eighth Seminar on Analysis of Algorithms, et dans son rôle d'éditeur dans le numéro spécial de *Combinatorics, Probability Computing* pour le "Eighth Seminar on Analysis of Algorithms".

À Jean-Christophe Dubacq pour m'avoir souvent dépanné informatiquement parlant et pour m'avoir inscrit à ens-lyon.org.

À Philippe Duchon pour de multiples services techniques rendus (notamment pour m'avoir prêté son portable lors d'un colloque Aléa), et pour avoir amorcé le travail sur le moment d'inertie des polyominos.

À Emmanuel Duits pour son intérêt pour feue ma page web sur la transmachinerie et pour m'avoir invité à son séminaire sur le paranormal, auquel je me suis rendu à titre purement académique, où j'ai d'ailleurs rencontré d'autres personnes du monde académique.

À Régis Dupont pour l'organisation du séminaire thésard du LIX.

À Marianne Durand pour m'avoir aidé à dénouer des points techniques mathématiques qui me bloquaient.

À l'École Normale Supérieure de Lyon pour la formation que j'y ai reçue, les personnes très enrichissantes que j'ai rencontrées et le salaire qui permet de s'affranchir des problèmes matériels.

À l'École Normale Supérieure de Paris dont j'ai assidument fréquenté les salles informatiques et où j'ai assisté à divers séminaires d'intérêt général, notamment ceux de l'atelier Internet et du laboratoire Pensée des Sciences, ainsi qu'à des cours de danse ou de langues. Grâce à elle j'ai pu acquérir un fragment de culture de l'honnête homme. J'y ai mis les pieds un temps comparable avec les écoles sous- et sus-citées avec lesquelles j'ai pourtant des liens plus officiels.

À l'École Polytechnique qui offre un excellent cadre de travail.

À Albert Einstein pour m'avoir fait comprendre que si je veux avoir la chance de faire une grande découverte un jour, je dois prendre des chemins de traverse. Mais je ne comprends peut-être rien à la recherche non plus.

À Nicole El Karoui pour m'avoir donné un exemplaire d'un de ses polys de l'X, et pour m'avoir accueilli dans son bureau pour une discussion scientifique informelle. À cette époque je ne m'intéressais pas spécialement aux mathématiques quantitatives et j'ignorais le rôle qu'elle jouait dans le monde financier.

À Andreas Enge qui fut mon co-bureau et avec qui nous avons eu d'intéressantes conversations, notamment sur la langue et la culture allemandes.

Aux étudiants que j'ai eus à Jussieu, l'X et l'IUT SRC de Marne-la-Vallée, qui m'ont beaucoup appris au niveau de la pédagogie, et à qui je souhaite bonne chance dans la vie, même s'il

y a peu de chance qu'ils lisent ces lignes (n'hésitez pas à me donner de vos nouvelles si c'est le cas !).

À Stéphane Fidanza pour m'avoir offert un exemplaire de sa thèse.

À Jim Fill pour son charisme lors des séminaires d'analyse d'algorithmes et des discussions toujours fort intéressantes et agréables, techniques ou autres.

À Philippe Flajolet pour sa disponibilité et sa bienveillance. Officieusement il a pratiquement joué le rôle d'un codirecteur de thèse.

À Cédric et Albane Flesselles pour leur amitié.

À Dominique Foata pour m'avoir envoyé une première version de la première partie de son livre portant notamment sur les martingales.

À Line Fonfrède pour sa compétence et sa rapidité d'esprit.

À Mireille Fouquet pour sa bonne humeur.

À mes deux frères pour m'avoir prêté leur salle de sport.

À FX Conseil pour avoir financé une partie de ma fin de thèse.

À Danièle Gardy pour m'avoir fait l'honneur de rapporter cette thèse et pour avoir relu des parties parfois rébarbatives de mes preuves.

À Stéphane Gaubert pour les plusieurs séances de travail que nous avons eus dans son bureau à l'INRIA, après des séminaires Algo, avant qu'il ne me voiture jusque dans mon quartier. J'étais allé le voir sur les conseils de Philippe Flajolet. Nous n'avons finalement pas eu de collaboration en commun mais ce n'est que partie remise.

À Florent Gillet pour m'avoir fait partager son savoir sur le calcul stochastique et Saint Seiya entre autres.

À Teresa Gomez-Diaz pour m'avoir ouvert illico presto un compte sur medicis, un certain dimanche.

À Dominique Grésillon pour son sérieux et son professionnalisme. Il m'a aussi voituré lors de Doctoriales de la gare vers le château de la Brosse-Montceaux et il m'a encouragé à poursuivre mes réflexions épistémologiques.

À Éric Guichard pour m'avoir initié à la sociologie de l'Internet et m'avoir fait graviter autour de l'ERST.

À Matthieu Guionnet pour ses compétences techniques en tant qu'administrateur système et responsable du vidéoprojecteur et des portables du laboratoire.

À Nicolas Gurel pour m'avoir voituré sur le trajet de mon retour à la maison.

À Patrice Hérault de l'IGM pour ses compétences informatiques et sa diligence.

À Elvira Horvath pour être à l'origine du Printemps Technologique et pour l'organiser de manière toujours très minutieuse, très professionnelle et très enthousiaste du Printemps Technologique, qui plus est dans une atmosphère sympathique mettant des doctorants en jeu. Merci également à toute l'équipe avec laquelle j'ai travaillée pour la préparation de la session 2003.

À Hsien-Kuei Hwang pour des conversations scientifiques très intéressantes et pour m'avoir donné les transparents de sa présentation sur la méthode des moments faite à Tatihou.

À Andrée Impérial pour son travail efficace à l'IGM.

À Jean Jacod pour m'avoir permis d'assister en auditeur libre à son cours sur les exponentielles de processus de Lévy et pour avoir par ailleurs répondu à quelques questions techniques sur les martingales.

À Philippe Jacquet pour avoir répondu à des questions techniques sur les tries et pour m'avoir prêté son DVD de Dune.

À Jean-Pierre Jouannaud pour son efficacité en tant que directeur du LIX et pour son intérêt pour mon devenir.

À Nevin Kapur pour les éléments de réponse qu'il m'a fournis concernant la viabilité et l'utilité d'un postdoc éventuel aux États-Unis.

À Sébastien Kunz-Jacques qui fut un collègue de bureau très enrichissant, pour des discussions autour de la théorie de Galois ou du corps des Télécom.

À Amaury Lambert pour m'avoir consacré quelques heures à répondre à quelques questions sur les probabilités continues.

À Dmitri Lebedev qui participa aussi à l'organisation du Printemps Technologique 2003.

Aux lecteurs de cette thèse.

Au LIX qu'il ne faut pas confondre avec l'X, pour le cadre de travail qu'il m'a fourni.

À Jean-Alex Losdat de la scolarité à l'ENS Lyon pour la qualité de son travail, et pour m'avoir aidé pour des problèmes administratifs notamment au début un peu houleux de ma thèse.

À Pierre Loiseleur pour m'avoir voituré jusqu'aux réunions du Printemps Technologique, et aussi parfois pour le retour de l'X à la maison.

À Guy Louchard pour m'avoir fait l'honneur de rapporter cette thèse et pour avoir fait partager ses connaissances sur le mouvement brownien.

À Jean-Gabriel Luque pour m'avoir aidé à préparer efficacement mes TD à Marne-la-Vallée, et pour des conseils éclairés que j'aurais peut-être dû parfois mieux suivre.

À Clémence Magnien pour l'organisation du séminaire des thésards du LIX.

À Hosam Mahmoud pour son théorème central limite sur les martingales et pour son intérêt pour mes travaux sur la simplification d'arbres.

À Jean-François Marckert pour avoir eu la patience de recompiler un de mes articles et m'avoir permis de faire quelques retouches de dernière minute.

À Conrado Martínez pour avoir accepté de faire partie de mon jury et de venir de Catalogne pour assister à ma soutenance. Il a fait montre d'une grande flexibilité par rapport aux dates de soutenance possibles alors que des soucis l'empêchaient de s'absenter longtemps de chez lui et en particulier de pouvoir faire du tourisme à Paris.

À Florence Marty pour son soutien et son écoute au début de ma thèse.

À Damien Massé pour m'avoir redonné accès au forum de l'ENS.

À Marianne Mathis qui a pu efficacement remplacer Andrée Impérial le cas échéant, pour les ordres de missions que j'ai pu lui faire pendant des périodes de vacances.

À MDP Conseil incarné par Marie-Dominique Pujol et Claude Forsans pour leur suivi post-doctoral.

À Ludovic Meunier pour ses carambars et son traducteur de Maple vers Latex, plus adapté à mes besoins que celui par défaut. Certaines formules sur des coefficients de matrice dans le chapitre sur les tries ont été écrites grâce à ce traducteur, sans lequel cela aurait été humainement très laborieux.

À moi-même pour avoir mené cette thèse à bout, malgré des conditions difficiles et parfois le sentiment d'être incompris par le haut et par le bas.

À Daniel Monniaux pour m'avoir ouvert un compte sur quatramaran et aidé à résoudre quelques problèmes techniques qui lui sont liés.

À Mioara Mügür Schachter pour les séminaires qu'elle organise dans le cadre du CeSEF et pour la documentation qu'elle m'a envoyée sur ses travaux.

À Michel Multan pour son efficacité dans le prêt inter-bibliothécaire et dans la recherche d'articles dans les bibliothèques souvent alentours mais parfois à l'étranger.

À ma soeur Myriam pour son dynamisme et son soutien moral et financier mais surtout moral.

À Markus Nebel pour d'intéressantes discussions et pour m'avoir à plusieurs reprises envoyé des exemplaires papiers de ses articles ou preprints portant sur le nombre de Strahler ou sur les ARN.

À Ralph Neininger pour avoir patiemment répondu aux multiples questions que je lui ai posées sur les théorèmes de point fixe et généralement pour son intérêt pour mes travaux.

À Basarab Nicolescu pour m'avoir replusieurs fois dans son bureau et expliqué plus en détail les concepts qu'il développe.

À Pierre Nicodème pour m'avoir fait inviter au séminaire de Bernard Mourrain à la Génopôle, et pour son organisation du séminaire bioinfo du LIX.

À Fabrice Nicot pour les Science & Vie Junior et autres passionnantes revues de la même maison qu'il a pu m'offrir, en particulier le hors-série consacré à Tintin.

À Roger Nifle pour son travail sur la trialectique et pour m'avoir contacté avant que je ne le contacte moi-même.

À Teodor Nitescu pour ses conseils éclairés au débutant en thèse que j'étais.

À Laurent Nottale pour ses séminaires pédagogiques et les documents qu'il a pu me transmettre sur ses travaux.

À mes parents sans l'aide matérielle de qui cette thèse n'aurait pas pu se faire.

À toute personne qui m'aidera à continuer à faire de la recherche par la suite. Comme *a priori* je pars dans le privé, je peux m'engager dans des voies de garage académiques mais scientifiquement porteuses et ambitieuses.

À Alain Plagne pour diverses conversations et pour quelques e-books scientifiques qu'il m'a communiqués.

À Dominique Poulalhon pour m'avoir indirectement permis d'accéder à la bibliothèque de mathématiques de la rue d'Ulm pendant un certain temps.

À Thai Quang pour d'intéressantes conversations sur les arbres et la bioinformatique notamment.

À Vlady Ravelomanana pour un service qu'il m'a subrepticement rendu, même si en fait je n'en ai pas tellement cherché à en tirer profit après.

À Évelyne Rayssac pour sa compétence et sa rapidité.

À Mireille Régnier pour des discussions techniques — séries génératrices, théorème de Perron-Frobenius, bio-informatique, etc. — et plus générales, et pour m'avoir fourni des références intéressantes.

À Maxime Rey pour m'avoir donné une idée de ce que l'on apprend dans les écoles d'ingénieurs informaticiens, et pour m'avoir posé des questions intéressante en algèbre ou en Maple.

À Philippe Robert pour avoir éclairé ma lanterne sur le th'eorème central limite sur les martingales et indiqué quelques références classiques.

À Émilia Robin pour m'avoir aidé à comprendre la situation des littéraires à Ulm.

À Uwe Rösler pour les diverses discussions techniques que nous avons eues lors des séminaires d'analyse d'algorithmes.

À Dominique Rossin pour sa compétence lors de la préparation d'enseignements que nous avons eus en commun, à Paris 7 et à l'X. Il m'a aussi voituré de temps en temps vers le séminaire d'Orsay.

À Xavier Sallantin pour m'avoir accueilli plusieurs fois à Béna et m'avoir communiqué de multiples documents par rapport à ses travaux.

À Michel Saloff-Coste pour son travail de recherche autour du management qui fait écho aux divers concepts que j'avais rencontrés en épistémologie.

À Bruno Salvy pour m'avoir consacré un peu de son temps et pour son intérêt pour les fonctions d'Airy et pour les équations différentielles.

À Gilles Schaeffer pour l'intérêt qu'il porte à mes travaux en général.

À Alexandre Sédoglavic pour m'avoir permis d'utiliser l'agrafeuse du Gage.

À ma soeur Séverine pour m'avoir donné les moyens de gagner quelques fractions de dioptries et pour m'avoir offert de quoi acheter une jolie raquette de tennis.

À Benoît Steyaert pour m'avoir à plusieurs reprises voituré sur le trajet de mon retour à la maison.

À Jean-Marc Steyaert pour m'avoir pris en thèse. Il m'a proposé un sujet de thèse très riche permettant de pouvoir définir ou élargir les problèmes que je voulais et aussi de choisir des questions de recherche en fonction des mathématiques que je voulais apprendre. Il m'a laissé toute la latitude et la liberté pour diriger mes recherches. Malgré un emploi du temps de ministre il a toujours su trouver du temps à me consacrer et m'a aidé à résoudre plusieurs points techniques cruciaux, sans parler de sa culture scientifique phénoménale. Il m'a trouvé des financements parallèles pour les périodes de vache maigre. Il m'a aussi souvent voituré sur le trajet de mon retour à la maison ou pour aller au séminaire d'Orsay ou autres. Je pense que Jean-Marc Steyaert a une énorme somme d'idées et qu'il a un potentiel encore inexploité.

À Wojciech Szpankowski pour avoir répondu à des questions techniques sur la dépoissonisation.

À Anne Tasso pour m'avoir formaté un cours de mathématiques correspondant au volume horaire que j'avais à enseigner en ATER.

À Emmanuel Thomé pour de multiples services informatiques rendus et pour son package Youpla qui m'a servi à ma soutenance de thèse.

À Nico Temme pour avoir répondu avec patience et compétence à mes questions sur les fonctions spéciales.

À Martine Thireau de l'IUT de l'Université de Marne-la-Vallée pour sa sympathie et son professionnalisme.

À Laurent Tuduri pour d'intéressantes discussions et pour quelques voiturages jusqu'à Lozère.

À l'Université Denis Diderot Paris 7 qui m'a accueilli pour mon monitorat, et où j'avais en fait déj'a passé mon DEUG dans des temps anciens.

À l'Université de Marne-la-Vallée qui m'a accueilli pour mon ATER et en particulier l'Institut Gaspard Monge très prolifique en recherche et dont je n'ai pas assez profité.

À Brigitte Vallée pour son rire éclatant.

À Ekaterina Vassilieva alias Katya qui fut mon co-bureau et avec qui nous avons eu d'intéressantes conversations, notamment sur les langues russe, française et japonaise.

À Jérôme Waldispühl pour sa sympathie, divers services rendus et sa collaboration pour le travail sur les ARN.

À Michel Weinfeld pour sa bienveillance et son efficacité lorsqu'il était directeur du LIX.

À Claude Weisbuch pour son travail autour des Doctoriales et pour m'avoir reçu dans son bureau concernant mes questions sur les postdocs aux États-Unis.

À Damien Wyart pour avoir installé Linux sur mon portable et m'avoir fait bénéficier de son savoir-faire informatique.

À Marc Yor pour m'avoir indiqué des références intéressantes sur les exponentielles d'excursions, et surtout m'avoir permis de m'inscrire à la bibliothèque de son Laboratoire de Probabilités à Paris 6. Il m'a aussi aidé dans l'écriture d'une de mes publications.

À Hervé Z. pour sa compétence et son suivi.

À toutes les personnes que j'aurais pu oublier de citer par manque de temps, parce qu'il faudra bien qu'un jour j'envoie ce manuscrit à l'impression.

Il va sans dire que les personnes que j'ai citées (ou non citées) m'ont en général rendu beaucoup plus de services que ceux que j'ai pu mentionner.

Table des matières

Introduction 1

Partie I Valuations liées aux expressions arithmétiques 5

1 Valuations arithmétiques sur des arbres binaires de Catalan 7

1.1 Arbres avec opérateurs "min","+" et opérandes dans un ensemble fini I d'entiers positifs . . . 8

1.1.1 "min" seulement . . . 8

1.1.2 "+" seulement . . . 8

1.1.3 "min" et "+" avec probabilités non nulles . . . 9

1.1.4 Interprétation probabiliste des a_k . . . 19

1.1.5 Un mot sur les arbres max, + . . . 19

1.2 Le paramètre 2-bud . . . 20

1.3 Arbres avec opérateurs "+" et "−" . . . 28

1.3.1 Arbres avec opérateurs "−" et opérandes dans l'ensemble fini I . . . 28

1.3.2 Arbres à opérateurs $''+'',''-''$ et opérandes dans un ensemble fini I . 32

1.4 Arbres +, × . . . 35

1.4.1 Séries génératrices des puissances des moments . . . 35

1.4.2 Asymptotique des moments . . . 36

1.4.3 Conséquences sur la distribution limite . . . 38

1.5 Conclusion . . . 39

2 Valuations arithmétiques sur des arbres tirés suivant la distribution des arbres binaires de recherche 41

2.1 Arbres +, − de recherche . . . 42

2.1.1 Moyenne et variance . . . 42

2.1.2 Point fixe . . . 45

2.1.3 Cas non gaussiens . . . 49

2.1.4 Perspectives : analyse de singularité . . . 53

2.2 Conclusion 54

3 Martingales **57**

3.1 Rappels sur le théorème central limite par martingales 57

3.2 Distribution des nombres de Narayana 58

3.2.1 Algorithme de Rémy 58

3.2.2 Les nombres de Narayana comme martingales 58

3.2.3 Distribution limite des nombres de Narayana par les martingales . . . 60

3.3 ABR $+,-$ et martingales 63

3.3.1 Moments de Y_n et S_n 63

3.4 Conclusion 70

4 Simplification et réduction d'arbres p-aires **71**

4.1 Introduction 71

4.2 Simplification d'arbres 72

4.2.1 Algorithme et fonctions génératrices 72

4.2.2 Une singularité algébrique 75

4.2.3 Loi limite 80

4.3 Lois p-aires nilpotentes 83

4.3.1 Algorithme et fonctions génératrices 83

4.3.2 Une singularité algébrique 84

4.3.3 Loi limite 90

4.4 Règles d'absorption 90

4.4.1 Séries génératrices univariées 91

4.4.2 Série génératrice bivariée 92

4.4.3 Distribution limite 93

4.5 Conclusion et questions ouvertes 95

5 Énergie de structures secondaires d'ARN **99**

5.1 Rappels 100

5.2 Grammaire 100

5.3 Énergie 101

5.4 Série génératrice (SG) 102

5.4.1 Nature de la loi limite 102

5.4.2 Moyenne et écart-type 103

5.4.3 Amélioration de la complexité de l'algo AMSAG ? 106

5.4.4 Figures 107

5.5 Conclusion 108

6 Lois stables pour arbres avec opérateurs "–" 111
6.1 Lois sans variance : préliminaires . . . 111
6.1.1 Loi de Zipf . . . 111
6.1.2 Densités Stables . . . 113
6.2 Loi limite stable . . . 114
6.2.1 Loi limite locale . . . 114
6.2.2 Limite faible . . . 117
6.3 Arbre – . . . 120
6.4 Conclusion . . . 121

Partie II Longueur de cheminement et moment d'inertie 123

7 Mouvement brownien 125
7.1 Rappels . . . 126
7.1.1 Propriété de scaling . . . 126
7.1.2 Formule de Feynman-Kac . . . 127
7.2 Pont brownien . . . 127
7.2.1 Aire . . . 129
7.2.2 Fonctionnelle quadratique . . . 129
7.2.3 Aire et moment d'inertie joints . . . 131
7.3 Excursion brownienne . . . 133
7.3.1 Aire sous l'excursion . . . 133
7.3.2 Fonctionnelle quadratique . . . 134
7.4 Calcul des moments . . . 136
7.4.1 Cas univarié . . . 136
7.4.2 Moments joints . . . 136
7.5 Conclusion . . . 138

8 Pompage de moments 139
8.1 Convergence des moments . . . 140
8.2 Aire et moment d'inertie joints sous les chemins de Grand Dyck . . . 141
8.2.1 Équation fonctionnelle . . . 141
8.2.2 Pompage des moments . . . 141
8.2.3 Moments joints de chemins de Grand Dyck . . . 144
8.2.4 La formule de Louchard pour les polynômes . . . 148
8.3 Arbres non orientés et polyominos parallélogrammes . . . 150
8.3.1 Arbres non orientés . . . 151

8.3.2 Polyominos 152
8.3.3 Des calculs plus simples 153
8.3.4 Encore d'autres objets ? 153
8.4 Méandre et mouvement brownien 153
8.4.1 Fonctions génératrices 154
8.4.2 Chemins de Dyck et chemins de Grand Dyck 154
8.4.3 Méandre et marche de Bernoulli 156
8.4.4 Aires signées 161
8.4.5 Autres fonctionnelles 162
8.5 Conclusion 164

9 Tries : occurrences de motifs et longueur de cheminement interne 165
9.1 Rappels 166
9.1.1 Zéros de $1 - p^{-s} - q^{-s}$ 167
9.1.2 Transformées de Mellin 169
9.1.3 Dépoissonisation 170
9.2 Occurrences de motifs 173
9.2.1 Moyenne et variance 175
9.2.2 Preuve résumée de la limite gaussienne par analyse complexe 178
9.2.3 Preuve de la limite gaussienne par méthode de point fixe 180
9.3 Longueur de cheminement interne 187
9.3.1 Dépoissonisation et matrice de covariance 192
9.3.2 Une limite gaussienne 193
9.3.3 Lemmes restants 200
9.4 Conclusion 203

Conclusion 205

Annexes 207

A Analyse de singularité 209
A.1 Analyse de singularité univariée 209
A.2 Limite discrète 210
A.3 Théorème des quasi-puissances 210
A.4 Remarques heuristiques 211
A.5 Moments factoriels 211

B Rappels sur le point fixe 213

Bibliographie **217**

Introduction

Les objets combinatoires de forme arborescente, ou arbres, se rencontrent partout en informatique [11, 194]. Ils peuvent représenter des structures de données pour augmenter l'efficacité des programmes, des expressions arithmétiques ou booléennes, des arbres de dérivation dans le cadre de la théorie des langages formels, et d'autres choses encore. Ce sont aussi des objets mathématiques à part entière, dans le cadre de la théorie des graphes. Un type d'arbre particulièrement étudié est celui des arbres binaires, constitués d'une racine et de nœuds comportant zéro ou deux sous-arbres binaires.

Des grandeurs associées à l'objet physique représenté par les arbres peuvent souvent se traduire de manière mathématique sur l'arbre par des paramètres récursifs qui peuvent s'étudier de manière analytique. Prenons par exemple le cas d'un arbre binaire complet représentant une expression arithmétique : ses feuilles ou nœuds externes comportent des opérandes, et les nœuds internes des opérateurs. Le nombre de registres pour effectuer le calcul de la valeur de cette expression, un registre comportant un opérande ou un opérateur, est égal au nombre de Strahler St défini de la manière suivante : le nombre de Strahler d'une feuille vaut 1, et le nombre de Strahler d'un arbre binaire de sous-arbres t_1 et t_2 vaut $\max(\mathrm{St}(t_1),\mathrm{St}(t_2))$ si $\mathrm{St}(t_1)$ et $\mathrm{St}(t_2)$ sont différents, et $1+\mathrm{St}(t_1)$ sinon.

Étudier un tel paramètre permet de savoir par exemple de quelle taille mémoire il faut disposer si l'on veut faire des calculs faisant intervenir un grand nombre d'opérations. Il n'est pas toujours possible de faire des calculs exacts, aussi essaie-t-on de trouver des valeurs asymptotiques pour les grandeurs d'intérêt. Dans le cas du nombre de Strahler, il s'agit par exemple de savoir, pour n opérations, de combien de registres on a besoin en moyenne, ou de connaître la distribution du nombre de Strahler de manière plus fine. À défaut d'identifier une distribution exacte de la variable aléatoire pour tout n, on essaie de connaître une distribution limite lorsqu'on fait tendre n vers l'infini pour pouvoir ensuite approcher la distribution de la variable aléatoire étudiée pour des n grands. Dans la pratique, n est déjà grand lorsqu'il est de l'ordre de la centaine. Le nombre de Strahler a été étudié en moyenne et en distribution lorsqu'on considère tous les arbres d'une taille donnée n (en nombre d'opérateurs) lorsque n tend vers l'infini.

Suivant les problèmes, les structures arborescentes qui sous-tendent les objets étudiés à taille n fixée ne suivent pas toujours la même distribution, et peuvent donc à ce titre soit être équiprobables, cas des arbres binaires de Catalan, soit ne pas l'être, comme par exemple pour les arbres de permutation, qui sont des arbres binaires étiquetés représentant des permutations. Un arbre de permutation peut représenter la succession d'opérations effectuées par un tri rapide (sur une instance différente de la permutation représentée par l'arbre). On parle alors d'arbre binaire de recherche (ABR). Le nombre de comparaisons effectuées, coût de l'algorithme, est alors égal à la longueur de cheminement de l'arbre, c'est-à-dire la somme des distances des nœuds à la racine.

Les arbres binaires de recherche sont plus équilibrés que les arbres binaires de Catalan, ce qui se traduit par des statistiques différentes pour des paramètres identiques. Le paramètre de Strahler, par exemple, a été étudié en moyenne sur des arbres binaires de recherche par Devroye,

et conduit à une moyenne différente de celle des arbres de Catalan. Une telle étude peut revêtir un intérêt plus mathématique que pratique, puisqu'étudier le nombre de Strahler sur des ABR a moins de sens que sur des arbres de Catalan.

Un autre type d'arbre abondamment étudié est le trie, dont le nom vient de retrieval. Il s'agit d'un arbre dictionnaire dans lequel on stocke n mots ou clés. On pourra en trouver une présentation dans le livre de Mahmoud [146] ou la thèse de Julien Clément [36]. Dans ce type d'arbre, contrairement aux arbres de Catalan ou de recherche, dans l'étude d'un paramètre donné, on ne conditionne pas par la taille mais par le nombre de clés, qu'on fait ensuite tendre vers l'infini.

La taille des tries admet une distribution limite gaussienne quand on fait tendre le nombre de clés vers l'infini, lorsqu'on engendre les lettres des clés par une source de Bernoulli [79, 78, 80, 113, 114, 146].

On peut sur ces exemples mesurer la grande diversité des situations et des méthodes d'analyse développées pour les étudier. L'étude de la distribution limite de paramètres d'arbre fait appel à des outils mathématiques aussi variés que sohistiqués.

La part belle est faite à l'analyse d'algorithmes, domaine initié par Knuth [129, 131, 132] à base de séries génératrices servant à énumérer les objets combinatoires. Manipuler des nombres exacts dans l'énumération d'objets est illusoire, il s'avère souvent nécessaire et utile de manipuler ces grandeurs de manière asymptotique, d'où l'intervention d'outils analytiques, notamment d'analyse complexe. La méthode de Darboux, développée d'abord par [9, 25, 39], permettait de connaître asymptotiquement certains nombres, comme par exemple le nombre de certains objets de taille n. Cette méthode fut raffinée par Flajolet et Odlyzko [75], à base de contours de Hankel, donnant lieu à une analyse de singularité précise des séries génératrices, permettant en plus de non seulement connaître l'asymptotique de ces nombres, mais d'obtenir des termes d'erreurs uniformes.

Dans le cadre de l'analyse de singularité bivariée, on retrouve l'omniprésence des lois gaussiennes à travers des clones du théorème central limite [8, 89]. Les termes d'erreurs uniformes ont permis une étude assez systématique de l'apparition de lois limites normales en distribution avec le théorème des quasi-puissances de Hwang [89, 108]. Des outils bien connus de l'analyse tels que la méthode du col et plus généralement d'intégration complexe permettent d'établir des lois limites locales dans le cas gaussien [95, 45, 96], voire dans le cas de lois de Rayleigh [49, 192] ou de lois stables dont en particulier la loi d'Airy-carte [6], ainsi surnommée car elle fait intervenir la fonction d'Airy et elle survient dans l'estimation de la taille de noyaux d'une grande variété de cartes [6, 10, 94].

Par la méthode de pompage des moments développée par Takács [198, 199, 200], on peut à partir d'équations fonctionnelles sur des séries génératrices bivariées, déduire des équations fonctionnelles univariées portant sur des moments de l'ordre qu'on veut, puis trouver les moments de la distribution limite. On peut quelquefois identifier alors la distribution limite comme dans les articles de Takács. On peut sinon au moins trouver des récurrences sur les moments permettant de tirer des informations sur la distribution limite [105]. Dans la plupart des cas, on peut au moins calculer les premiers moments, espérance et variance en particulier, quitte ensuite à trouver la distribution limite par des méthodes différentes (martingales, point fixe...).

En dehors de ces méthodes générales, des moyens ad hoc construits sur l'analyse de singularité, en rapport avec le problème considéré, nécessitant finesse et créativité, permettent de trouver la distribution limite, principalement dans le cas des arbres de Catalan. C'est souvent le cas pour des paramètres extrémaux mettant en jeu l'opérateur max de type nombre de Strahler [77, 76, 93, 124, 125, 161, 164, 165, 162] ou hauteur d'arbre [74, 71].

Certains types d'arbre comme les tries nécessitent l'usage d'outils analytiques particulièrement sophistiqués, comme la transformation de Mellin [72, 78, 83, 84, 146, 195, 211], et le cycle poissonisation-dépoissonisation [78, 195]. Dans le chapitre 9, nous nous intéressons à des tries construits sur des mots engendrés par une source de Bernoulli. Un premier paramètre d'intérêt est le nombre d'occurrences d'un motif (trie) donné dans les tries avec une source de Bernoulli, Un autre paramètre d'intérêt est la longueur de cheminement interne des tries, dans la même lignée que la taille (nombre de nœuds internes) des tries, déjà étudiée [78, 113, 114].

Des outils classiques de probabilités tels que les martingales permettent d'obtenir l'existence de loi limite avec des convergences assez fortes, comme dans L_1 ou dans L_2. Ainsi, le coût du tri rapide, autrement dit la longueur de cheminement interne des arbres binaires de recherche fait l'objet d'une étude à base de martingales [181], prolongée pour le polynôme de niveau des arbres binaires de recherche par [30].

Si l'utilisation des martingales dans le cadre des convergences fortes ne permet pas de déterminer les distributions limites, il existe, avec alors des convergences faibles, un théorème central limite par martingales [54, 53, 104, 112], utilisé dans [145, 191] dans des problèmes proches des modèles d'urnes.

Si ce théorème peut s'avérer assez puissant, comme en témoigne l'étude assez générale de fonctions à péage dans les tries faite dans [187], un outil au moins assez puissant et pratique se trouve dans les méthodes de point fixe, présentées dans l'annexe B, particulièrement bien adaptées aux arbres binaires de recherche et aux tries [109, 169, 178, 179, 180, 201]. Notons que ces méthodes s'appliquent assez mal aux arbres de Catalan, car dans ces arbres la probabilité d'avoir un sous-arbre de taille 1 vaut 1/2. Seul un article de Svante Janson [115] parvient à étudier un paramètre sur des arbres de Catalan. Les techniques utilisées dans cet article ne peuvent s'appliquer qu'à des variables ne nécessitant aucune normalisation, et donc restent difficiles à étendre.

Mentionnons enfin que le mouvement brownien, objet limite des marches aléatoires, qui a connu un grand développement au cours du XXe siècle, connaît une littérature abondante, avec un outillage analytique sophistiqué en termes de probabilités continues. Voir par exemple [62, 63, 100, 111, 122, 123, 139, 185]. Le contour des arbres de Catalan, ou plus généralement d'arbres simplement engendrés, converge faiblement vers une excursion brownienne [102, 151, 152], si bien que l'on peut récupérer des résultats connus sur la distribution de fonctionnelles de l'excursion brownienne et les interpréter comme distribution limite [13, 110, 128, 140, 141, 142, 143, 183, 213, 214] de certains paramètres d'arbre. (Une fonctionnelle d'un mouvement brownien est une variable aléatoire construite sur une trajectoire tirée au hasard).

Le paysage semble donc assez hétérogène. Nous avons essayé dans ce mémoire d'en donner un aperçu plus linéaire et plus structuré.

Le corps de cette thèse se divise en deux parties. La première est associée à l'étude des paramètres d'arbre associés aux valuations arithmétiques. La seconde étudie des paramètres du type longueur de cheminement interne (somme des distances des nœuds à la racine) ou moment d'inertie (somme des carrés des distances des nœuds à la racine).

La première partie comporte les chapitres 1 à 6.

Dans le **chapitre 1**, on définit la valuation d'un arbre d'expression arithmétique comme le résultat de cette expression arithmétique. On y étudie la distribution limite de valuations arithmétiques sur des arbres binaires de Catalan, avec opérandes $\{\min, +\}$, $\{+, -\}$, et $\{+, \times\}$. Il est fait usage d'analyse de singularité, pour prouver une limite discrète dans le cas des opérandes $\{\min, +\}$, une limite gaussienne dans le cas des opérandes $\{+, -\}$, et pour calculer les moments dans le cas des opérandes $\{+, \times\}$.

Le **chapitre 2** étudie la distribution limite de valuations arithmétiques sur des arbres binaires de recherche, avec opérandes $\{+,-\}$. L'utilisation de fonctions génératrices et d'analyse de singularité sert surtout à déterminer l'espérance et la variance de la valuation. On établit la nature de la loi limite — gaussienne dans la plupart des cas — à l'aide de méthodes de point fixe.

Le **chapitre 3** retrouve une partie des résultats précédents en utilisant le théorème central limite sur les martingales. On établit à nouveau le caractère gaussien de la distribution limite des nombres de Narayana d'une part, et des valuations arithmétiques sur des arbres binaires de recherche d'autre part, avec opérandes $\{+,-\}$ lorsque l'opérateur $+$ apparaît avec probabilité moindre que celle de l'opérateur $-$.

Le **chapitre 4** étudie la distribution limite de la taille d'arbres réduits, où la réduction d'un arbre d'expression logique ou arithmétique suit une règle logique ou arithmétique, comme $X \wedge X \to X$. Il est fait usage de fonctions génératrices et d'analyse de singularité. La principale difficulté consiste à établir des inégalités montrant qu'on se trouve dans les conditions du théorème des quasi-puissances.

Le **chapitre 5** étudie la distribution de l'énergie des molécules d'ARN et ébauche comment améliorer en complexité un algorithme permettant de trouver une structure secondaire minimisant l'énergie d'un mot d'ARN donné.

Le **chapitre 6** étudie les valuations arithmétiques sur des arbres binaires de Catalan, avec opérandes $-$ et certaines lois sans variances aux nœuds externes, en faisant appel à de l'analyse de singularité. On y ébauche un début de généralisation du théorème des quasi-puissances, en faisant cette fois-ci apparaître des lois stables à la limite.

La seconde partie comporte les chapitres 7 à 9.

Le **chapitre 7** étudie les fonctionnelles de type aire ou quadratique du pont brownien et de l'excursion, à l'aide de la formule de Feynman-Kac. Ces résultats sont déjà connus, mais nous donnons une nouvelle preuve de la distribution de l'intégrale du carré de l'excursion, à l'aide du théorème de Louchard, et nous faisons une ébauche de l'étude de la distribution jointe entre les deux fonctionnelles. Pour cette dernière nous donnons la double transformée de Laplace de la distribution jointe, et effectuons le calcul du moment joint d'ordre $(1,1)$.

Le **chapitre 8** reprend cette étude pour le pont brownien, l'excursion, les méandres et les marches de Bernoulli, en utilisant de l'analyse de singularité et du pompage de moments. Nous montrons comment calculer par récurrence les moments joints d'ordre arbitraire, qui seraient plus difficilement accessibles en utilisant les objets continus exposés dans le chapitre précédent.

Le **chapitre 9** étudie la distribution limite du nombre d'occurrences d'un certain motif dans les tries, et de la longueur de cheminement interne des tries, avec une source de Bernoulli. Pour le premier paramètre, nous faisons une première preuve en utilisant uniquement des outils de variable complexe, puis une seconde preuve utilisant des méthodes de point fixe. Pour le second, la seule utilisation de l'analyse complexe semble insuffisante, et nous devons avoir recours aux méthodes de point fixe.

L'**annexe A** rappelle des résultats fondamentaux de l'analyse de singularité.

L'**annexe B** rappelle les résultats sur les méthodes de point fixe, utilisés dans le chapitre 2 et surtout le chapitre 9.

Première partie

Valuations liées aux expressions arithmétiques

Chapitre 1

Valuations arithmétiques sur des arbres binaires de Catalan

Dans ce chapitre, nous étudions la distribution limite du résultat d'expressions arithmétiques tirées au hasard, dont on fait tendre la taille (le nombre d'opérateurs) vers l'infini. À une expression arithmétique peut être associé un arbre d'expression. Nous appellerons valuation d'un arbre d'expression le résultat de l'expression arithmétique correspondante.

Nous tirons les expressions arithmétiques de la manière suivante. Nous nous fixons un entier naturel n, nous tirons uniformément au hasard un arbre binaire complet de Catalan parmi les arbres binaires complets de Catalan de taille n, *i.e.* à nœuds internes. Puis nous mettons aux nœuds externes, de manière indépendante, un entier choisi parmi un nombre fini d'entiers formant un ensemble I, suivant une probabilité fixée quelconque : à chaque nœud externe, l'entier i apparaît avec la probabilité f_i. Nous plaçons enfin des opérateurs choisis dans un ensemble de deux opérateurs, par exemple $\{+,-\}$, avec probabilités respectives α et β. Nous noterons $\mathcal{T}((+,\alpha),(-,\beta),I)$ l'ensemble de ces expressions arithmétiques, pondérées par le produit des probabilités des opérateurs et opérandes présents dans l'expression. Ainsi, si l'arbre d'expression t possède $U(t)$ opérateurs "+", $|t|$ noeuds internes et $N_i(t)$ opérandes i, alors son poids est donné par l'expression

$$\text{poids}(t) = \alpha^{U(t)}\beta^{|t|-U(t)}\prod_{i\in I} f_i^{N_i(t)}.$$

Le paramètre étudié ici est la valeur de l'expression arithmétique tirée au hasard. Tout au long de ce chapitre, nous noterons X_n la variable aléatoire égale à cette valeur, pour une expression arithmétiques de taille n. Notre but consiste à trouver la distribution limite de X_n lorsque n tend vers l'infini, en terme de convergence en distribution.

Nous nous intéresserons à des jeux d'opérateurs $\{\min,+\}$, $\{+,-\}$ et $\{+,\times\}$. Ce dernier jeu d'opérateurs apparaît dans [12]. Nous ferons au passage l'étude du paramètre 2-bud, introduit par Blondel [16], qui suit la même méthodologie que l'étude du jeu d'opérateurs $\{\min,+\}$.

Ces résultats pour l'instant sont seulement intéressants en eux-mêmes, et il n'en existe pas d'application directe. Il faut remarquer que l'aléa provient d'une grande part de la forme des arbre, si bien que les problèmes ici sont très différents de ceux qu'étudient Devroye et Kamoun dans [42], où tous les arbres sont parfaits.

Dans ce chapitre est fait grand usage de séries génératrices et d'analyse de singularité. Le lecteur non familiarisé avec ces outils pourra consulter avec profit [81, 99, 188, 193] pour la partie combinatoire, et [81, 85] pour la partie analytique.

1.1 Arbres avec opérateurs "min","+" et opérandes dans un ensemble fini I d'entiers positifs

Nous mélangeons des opérateurs minimum et addition. Pour cela, nous définissons, sur tout arbre t, la valuation v par

$$v(t) = \begin{cases} k & \text{si } t = \boxed{k} \text{ avec } k \in I, \\ \min(v(t_1), v(t_2)) & \text{si } t = \begin{array}{ccc} & \min & \\ / & & \backslash \\ t_1 & & t_2 \end{array}, \\ v(t_1) + v(t_2) & \text{si } t = \begin{array}{ccc} & + & \\ / & & \backslash \\ t_1 & & t_2 \end{array}. \end{cases}$$

Nous tirons chaque entier k avec la probabilité f_k aux nœuds externes, indépendamment les uns des autres. On note $F(u) = \sum_{k \in I} f_k u^k$ la fonction génératrice des entiers aux nœuds externes. À chaque nœud interne, indépendamment des autres nœuds, nous mettons un "+" avec probabilité α et un "min" avec probabilité $\beta = 1 - \alpha$.

Il est évident que la valuation d'expressions mettant uniquement en jeu des "min" convergera en distribution vers une masse de Dirac, et que celle qui met uniquement en jeu des additions convergera en distribution — après normalisation — vers une loi gaussienne. Nous prouvons que mélanger des "min" et des "+" mène à des lois limites discrètes. Nous renvoyons le lecteur à l'annexe A pour une définition de ce qu'est une limite discrète. Notons qu'on parle alors de convergence en distribution.

Nous allons donc d'abord traiter les cas extrêmes $\alpha = 0$ et $\beta = 0$, qui sont les plus faciles. Nous traitons ensuite le cas quelconque $\alpha \cdot \beta \neq 0$.

Nous notons X_n la variable aléatoire égale à la valuation d'un arbre de taille n tiré au hasard dans $\mathcal{T}((+,\alpha),(-,\beta),I)$.

1.1.1 "min" seulement

Soient $k_1 < k_2 < \cdots < k_{|I|}$ les éléments ordonnés de I. Si on n'a que des min, la structure d'arbre n'intervient visiblement pas : calculer la valuation des expressions revient à calculer le mininum des opérandes. Pour tout i, la probabilité que la valuation X_n d'une expression arithmétique à n opérandes vaille k_i est donc donnée par

$$\mathbf{Pr}[X_n = k_i] = \left(\sum_{k \geqslant k_i} f_k\right)^{n+1} - \left(\sum_{k > k_i} f_k\right)^{n+1},$$

et la loi limite est une masse de Dirac en $k = k_1 = \min I$. En d'autres termes, X_n converge en distribution vers la variable aléatoire X définie par $\mathbf{Pr}[X = k_1] = 1$ et $\mathbf{Pr}[X = k_i] = 0$ si $k_i \neq k_1$.

1.1.2 "+" seulement

Lorsqu'on ne met que des opérateurs "+", c'est comme si on ne faisait qu'additionner des variables aléatoires identiques et indépendamment distribuées selon la probabilité de série génératrice F. On retrouve donc le théorème central limite.

Proposition 1.1 *Avec les notations ci-dessus, si les opérateurs sont tous égaux à l'addition, et si I n'est pas réduit à un singleton, alors $\frac{X_n-(n+1)F'(1)}{\sqrt{(n+1)\,\mathbf{Var}(F)}}$ converge en loi vers $\mathcal{N}(0,1)$ lorsque n tend vers l'infini.*

Démonstration. La structure de l'arbre n'intervient pas, et la valuation de l'arbre de taille n admet F^{n+1} pour fonction génératrice de probabilité. Le caractère gaussien de la limite est un résultat classique qu'on peut trouver en calculant la fonction caractéristique de X_n :

$$\begin{aligned}\log\phi_{X_{n-1}}(t) &= n\log\Big(\frac{1}{|I|}\sum_{k\in I}e^{ikt}\Big)=n\log\Big(1+\frac{1}{|I|}\sum_{k\in I}\sum_{p\geqslant 1}\frac{(itk)^p}{p!}\Big)\\ &= n\Big(\frac{1}{|I|}\sum_{k\in I}ikt-\frac{1}{|I|}\sum_{k\in I}\frac{k^2}{2}t^2+\frac{1}{2}\frac{1}{|I|^2}\Big(\sum_{k\in I}k\Big)^2t^2+O(t^3)\Big).\end{aligned}$$

Posant

$$X_n^*=\frac{X_n-(n+1)\frac{1}{|I|}\sum_{k\in I}k}{\sqrt{n+1}\left(\frac{1}{|I|}\sum_{k\in I}k^2-\frac{1}{|I|^2}\left(\sum_{k\in I}k\right)^2\right)},$$

nous avons $\phi_{X_n^*}(t)=e^{-t^2/2+O(t^3/\sqrt{n+1})}$ qui pour tout t tend vers $e^{-t^2/2}$ lorsque n tend vers l'infini, et cela établit la convergence en distribution vers une limite gaussienne. □

1.1.3 "min" et "+" avec probabilités non nulles

Nous tirons maintenant "+" avec probabilité α et "min" avec probabilité $\beta=1-\alpha$, en supposant $\alpha\beta\neq 0$. Les principaux résultats que nous allons montrer dans ce paragraphe consistent d'une part en le théorème :

Théorème 1.1 *Lorsque n tend vers l'infini, la distribution de la valuation X_n converge en loi vers une limite discrète X donnée par*

$$\mathbf{Pr}[X=k]=b_k,$$

où la suite (b_k) est définie conjointement avec une suite (a_k) par :

$$\begin{aligned}a_0 &= \frac{1}{2\alpha-1}\left(\alpha-\sqrt{\alpha^2-f_0(2\alpha-1)}\right)\ \text{si}\ \alpha\neq\frac{1}{2},\ a_0=f_0\ \text{sinon},\\ b_0 &= \frac{1-\alpha}{2\alpha-1}\left(\frac{\alpha}{\sqrt{\alpha^2-f_0(2\alpha-1)}}-1\right)\ \text{si}\ \alpha\neq\frac{1}{2},\ b_0=f_0\ \text{sinon},\end{aligned}$$

et, pour $k\geqslant 1$,

$$\left\{\begin{aligned}a_k &= \frac{1}{\beta}\sqrt{\left[1-\alpha a_0-\beta\left(1-\sum_{j=0}^{k-1}a_j\right)\right]^2+\beta\left(f_k+\alpha\sum_{j=1}^{k-1}a_ja_{k-j}\right)},\\ &\quad -\frac{1}{\beta}\left[1-\alpha a_0-\beta\left(1-\sum_{j=0}^{k-1}a_j\right)\right]\\ b_k &= \frac{\alpha\sum_{j=0}^{k-1}b_ja_{k-j}-\beta a_k\left(\sum_{j=0}^{k-1}b_j-1\right)}{1-\alpha a_0-\beta\left(1-\sum_{j=0}^{k}a_j\right)}.\end{aligned}\right.$$

et d'autre part en le théoréme suivant, établissant l'asymptotique des b_k :

Théorème 1.2 *Il existe des constantes C et ρ_B, dépendant de α, telle que, pour les b_k non nuls,*

$$b_k \underset{k\to\infty}{\sim} C\frac{\rho_B^{-k}}{\sqrt{k}}.$$

Remarque. ρ_B n'est rien d'autre que le rayon de convergence de la série $\sum_k b_k z^k$.

Avant de donner la preuve proprement dite des théorèmes 1.1 et 1.2, nous allons établir des résultats préliminaires sous forme de lemmes.

Définissons, pour $k \geqslant 0$, les fonctions génératrices des arbres de valuation k

$$W_k(z) = \sum_{T, v(T)=k} z^{|T|},$$

et notons $f_k = \mathbf{Pr}[X_0 = k] = [u^k]F(u)$ les probabilités aux feuilles. Nous allons d'abord montrer que tous les W_k admettent une singularité dominante égale à $1/4$ (tant que $W_k \not\equiv 0$), et satisfont un développement asymptotique similaire.

Lemme 1.1 *Tout $W_k \not\equiv 0$ possède un développement asymptotique au voisinage de $1/4$ (sa singularité dominante) de la forme*

$$W_k = 2a_k - 2b_k\sqrt{1-4z} + O(1-4z), \text{ avec } a_k > 0,\ b_k \geqslant 0.$$

Démonstration. Écrivons $W_{\leqslant k} = \sum_{p\leqslant k} W_p$, et utilisons les notations similaires $W_{<k}$, $W_{\geqslant k}$, $W_{>k}$. Notons $T(z) = W(z,1) = 1 + zT(z)^2 = \frac{1-\sqrt{1-4z}}{2z}$. Un arbre est de valuation 0 s'il s'agit d'un arbre réduit à l'opérande 0, ou bien d'un arbre avec opérateur + ou min à la racine et deux sous-arbres de valuation 0, ou bien d'un arbre avec opérateur min à la racine et deux sous-arbres dont l'un (le sous-arbre gauche ou bien le sous-arbre droit) est de valuation 0, et l'autre de valuation strictement plus grande que 0. Rappelons en outre que l'opérateur + est pondéré par α et l'opérateur min par $\beta = 1-\alpha$. De cette décomposition on déduit

$$\begin{aligned} W_0 &= f_0 + (\alpha+\beta)zW_0^2 + 2\beta z W_0 W_{>0} \\ &= f_0 + zW_0^2 + 2\beta z W_0(T - W_0) \\ &= f_0 + (1-2\beta)zW_0^2 + 2\beta z W_0 T. \end{aligned}$$

Si $\alpha \neq \beta$, W_0 vérifie

$$\begin{aligned} W_0 &= \frac{(1-2\beta zT) - \sqrt{(1-2\beta zT)^2 - 4f_0(1-2\beta)z}}{2(1-2\beta)z} \\ &= 2a_0 - 2b_0\sqrt{1-4z} + O(1-4z), \end{aligned}$$

avec $a_0 = \frac{1}{2\alpha-1}\left(\alpha - \sqrt{\alpha^2 - f_0(2\alpha-1)}\right)$ et $b_0 = \frac{1-\alpha}{2\alpha-1}\left(\frac{\alpha}{\sqrt{\alpha^2-f_0(2\alpha-1)}} - 1\right)$.

Si $\alpha = \beta = 1/2$, alors

$$W_0 = f_0 + zW_0^2 + zW_0W_{>0} = f_0 + zW_0^2 + zW_0(T - W_0) = f_0 T.$$

Donc $a_0 = b_0 = f_0$, et on peut remarquer que les expressions de $a_0(\alpha)$ et $b_0(\alpha)$ trouvées ci-dessus pour $\alpha \neq 1/2$ tendent vers f_0 lorsque α tend vers $1/2$.

En distinguant, pour tout k, parmi les arbres de valuation k, les arbres réduits à une feuille, les arbres avec opérateur + à la racine, et les arbres avec opérateur min à la racine, on trouve facilement que les W_k pour $k \geqslant 1$ sont donnés par la récurrence

$$\begin{aligned} W_k &= f_k + \alpha z \sum_{l=0}^{k} W_l W_{k-l} + 2\beta z W_{>k} W_k + \beta z W_k^2 \\ &= f_k + \alpha z \sum_{l=0}^{k} W_l W_{k-l} + 2\beta z (T - W_{\leqslant k}) W_k + \beta z W_k^2 \\ &= f_k + \alpha z \sum_{l=1}^{k-1} W_l W_{k-l} + 2z\big(\alpha W_0 + \beta(T - W_{\leqslant k-1})\big) W_k - \beta z W_k^2 \qquad (1.1) \\ &= \frac{-(1 + 2z(\beta(W_{\leqslant k-1} - T) - \alpha W_0))}{2\beta z} \\ &\quad + \frac{\sqrt{(1 + 2z(\beta(W_{\leqslant k-1} - T) - \alpha W_0))^2 + 4\beta z \left(f_k + \alpha z \sum_{l=1}^{k-1} W_l W_{k-l}\right)}}{2\beta z}. \end{aligned}$$

W_k correspond à la solution de l'équation quadratique analytique en zéro, ayant des coefficients positifs. Le comportement autour de la singularité $\frac{1}{4}$ est fondamentalement identique pour tous les W_k non nuls :

$$W_k(z) = 2a_k - 2b_k\sqrt{1 - 4z} + O(1 - 4z).$$

Les W_k sont des séries avec des coefficients positifs majorés par C_n, alors les W_k non nuls sont croissants sur $[0, 1/4]$ avec $2a_k = \lim_{z \to \frac{1}{4}} W_k(z) > 0$; les séries correspondantes $2a_k - W_k$ sont des fonctions décroissantes et positives sur $[0, 1/4[$, et sont donc positives près de $1/4$ avec $b_k \geqslant 0$. Finalement, les W_k non nuls vérifient

$$W_k = 2a_k - 2b_k\sqrt{1 - 4z} + O(1 - 4z), \text{ avec } 2a_k > 0,\ 2b_k \geqslant 0.$$

□

On peut déjà deviner que les (b_k) constituent la distribution limite de X_n. Pour le montrer rigoureusement, nous devons montrer que les sommes $\sum_k a_k$ et $\sum_k b_k$ convergent et sont égales à 1. Nous allons en fait établir des propriétés un peu plus fortes, permettant d'obtenir l'asymptotique de la suite (b_k).

Examinons maintenant les propriétés des suites (a_k) et (b_k). Nous pouvons calculer chaque a_k et b_k à l'aide d'une récurrence utilisant les termes précédents des deux suites :

Lemme 1.2 *Pour $k \geqslant 1$, les suites (a_k) et (b_k) sont données par la récurrence :*

$$\begin{cases} a_k = \frac{1}{\beta}\sqrt{\left[1 - \alpha a_0 - \beta\left(1 - \sum_{j=0}^{k-1} a_j\right)\right]^2 + \beta\left(f_k + \alpha \sum_{j=1}^{k-1} a_j a_{k-j}\right)}, \\ \qquad -\frac{1}{\beta}\left[1 - \alpha a_0 - \beta\left(1 - \sum_{j=0}^{k-1} a_j\right)\right] \\ b_k = \frac{\alpha \sum_{j=0}^{k-1} b_j a_{k-j} - \beta a_k \left(\sum_{j=0}^{k-1} b_j - 1\right)}{1 - \alpha a_0 - \beta\left(1 - \sum_{j=0}^{k} a_j\right)}. \end{cases}$$

Démonstration. Dans l'équation (1.1), on développe les séries en $x = \sqrt{1-4z}$ et on identifie les termes en x^0 et x^1. Nous obtenons les relations

$$2a_k = f_k + \alpha \sum_{j=1}^{k-1} a_j a_{k-j} + 2\beta a_k (1 - \sum_{j=0}^{k-1} a_j) + 2\alpha a_0 a_k - \beta a_k^2, \tag{1.2a}$$

$$-2b_k = -2\alpha \sum_{j=0}^{k} b_j a_{k-j} - 2\beta a_k b_k + 2\beta a_k (\sum_{j=0}^{k} b_j - 1) - 2\beta b_k (1 - \sum_{j=0}^{k} a_j). \tag{1.2b}$$

Nous résolvons ces équations en choisissant la même branche de Riemann que pour les W_k, et nous trouvons les relations de récurrence voulues. □

Lemme 1.3 $\sum_{k\geqslant 0} a_k = 1$.

Démonstration. Comme nous travaillons avec des séries dont les coefficients sont positifs, nous avons l'égalité $W(1/4,1) = \lim_{z\to 1/4} W(z,1) = \lim_{z\to 1/4} \sum_{k\geqslant 0}^{\infty} W_k(z) = \sum_{k\geqslant 0}^{\infty} \lim_{z\to 1/4} W_k(z)$ $= \sum_{k\geqslant 0}^{\infty} W_k(1/4)$, ce qui signifie que la somme $\sum_{k\geqslant 0} 2a_k$ est convergente de limite 2, c'est-à-dire $\sum_{k\geqslant 0} a_k = 1$. □

Lemme 1.4 *Le rayon de convergence R_A de la série A définie par $A(z) = \sum_k a_k z^k$ est strictement plus grand que 1.*

Démonstration. Pour montrer que le rayon de convergence de A est strictement plus grand que 1, nous allons introduire une série entière dont les coefficients sont plus grands que ceux de A, et dont nous pouvons montrer plus facilement que le rayon de convergence est strictement plus grand que 1.

Construire une telle série n'est pas trivial, aussi introduisons une famille de séries $M^{(k,n)}$ indexée par un double paramètre (k,n), où k et n sont entiers avec $k \geqslant 2$. Précisons comment s'interprètent les entiers n et k.

Si on écrit $M^{(k,n)} = \sum_{j\geqslant 1} m_j^{(k,n)} z^j$, alors les premiers coefficients de $M^{(k,n)}$ sont identiques à ceux de A jusqu'au rang $n-1$, et diffèrent de ceux de A à partir du rang n. Autrement dit, on a $m_j^{(k,n)} = a_j$ pour $j \leqslant n-1$.

L'entier k varie avec la proximité de la définition par récurrence des termes de $M^{(k,n)}$ par rapport à la définition par récurrence des termes de A, à partir du rang n. Plus précisément, pour $j \geqslant n$, on pose

$$m_j^{(k,n)} = \theta_k \sum_{q=1}^{j-1} m_q^{(k,n)} m_{j-q}^{(k,n)},$$

où

$$\theta_k = \frac{\alpha}{2\left(\alpha + (\beta-\alpha)a_0 + \beta \sum_{j=1}^{k-1} a_j\right)}.$$

Utilisant l'équation (1.2) et les inégalités $1/\theta_{k'} \geqslant 1/\theta_k$ pour tous $k < k'$, et $\sqrt{1+x} \leqslant 1 + \frac{1}{2}x$ pour tout $x \geqslant 0$, nous avons, pour $j \geqslant k$, $a_j \leqslant \theta_k \sum_{q=1}^{j-1} a_q a_{j-q}$. Par conséquent, pour tout $j \geqslant 1$, $a_j \leqslant m_j^{(k,n)}$ et, pour établir que $R_A > 1$, il suffit de montrer qu'il existe k, n tels que $R_{M^{(k,n)}} > 1$. Fixons maintenant K tel que $\theta_K < 1/2$, *e.g.* $K = \min I$.

Pour toute série $H = \sum_{n\geqslant 0} h_n z^n$ et tous entiers $0 < p < q$, introduisons la notation $H_{p..q} = \sum_{n=p}^{q} h_n z^n$. La définition de $M^{(K,n)}$ peut se réécrire de la manière suivante :

$$\begin{cases} M^{(K,n)}_{1..n-1} = \sum_{j=1}^{n-1} m_j^{(K,n)} z^j = \sum_{j=1}^{n-1} a_j z^j, \\ M^{(K,n)} - M^{(K,n)}_{1..n-1} = \theta_k \left((M^{(K,n)})^2 - ((M^{(K,n)})^2)_{1..n-1} \right). \end{cases}$$

Ainsi, $M^{(K,n)}$ est défini par

$$\begin{aligned} 2\theta_k M^{(K,n)} &= 1 - \sqrt{1 - 4\theta_k (M_{1..n-1} - \theta_k (M^2)_{1..n-1})} \\ &= 1 - \sqrt{1 - 4\theta_k (\theta_k A_{1..n-1} - ((A-a_0)^2)_{1..n-1})} \\ &= 1 - \sqrt{((1 - 2\theta_k (A - a_0))^2)_{0..n-1}}. \end{aligned}$$

$G^{(K)} = (1 - 2\theta_k(A - a_0))^2$ est une série positive sur [0,1] telle que $\sum_{j\geqslant 0} g_j^{(K)}$ est absolument convergente car chaque $\left| g_j^{(K)} \right|$ est majoré par la quantité $[z^j](1 + 2\theta_k(A - a_0))^2(z)$. Donc $\sum_{j\neq 0} g_j^{(K)} z^j$ converge uniformément sur [0,1], et il existe N tel que $G^{(K)}_{0..N-1}$ est positif sur [0,1]. Pour conclure, $M^{(K,N)}$ possède un rayon de convergence $R_{M^{(K,N)}}$ strictement plus grand que 1, et finalement $R_A \geqslant R_{M^{(K,N)}} > 1$. □

Lemme 1.5 $\sum_{k\geqslant 0} b_k \leqslant 1$.

Démonstration. Introduisons les séries $Y_k(z) = \frac{a_k - 2zW_k(z)}{\sqrt{1-4z}}$ qui sont analytiques sur $[0, 1/4[$, et dont la somme pour $z < \frac{1}{4}$ vaut

$$\sum_k Y_k(z) = \sum_k \frac{a_k - 2zW_k(z)}{\sqrt{1-4z}} = \frac{\sum_k a_k - 2z\sum_k W_k(z)}{\sqrt{1-4z}} = \frac{1 - 2zW(z)}{\sqrt{1-4z}} = \frac{\sqrt{1-4z}}{\sqrt{1-4z}} = 1.$$

Les Y_k sont positifs sur $[0, 1/4[$ et vérifient $Y_k(z) = b_k + O(\sqrt{1-4z})$. Pour tous $n \in \mathbb{N}$ et $z < \frac{1}{4}$, nous avons $\sum_{k\leqslant n} Y_k(z) \leqslant 1$, et en faisant tendre z vers $1/4$, nous obtenons $\sum_{k\leqslant n} b_k \leqslant 1$. Comme les b_k sont positifs, la somme $\sum_{k\geqslant 0} b_k$ est convergente, de somme $B(1)$ plus petite que 1 (au sens large). □

Lemme 1.6 *La série $B(z) = \sum_{k\geqslant 0} b_k z^k$ possède un rayon de convergence au moins égal à celui de la série $A(z) = \sum_{k\geqslant 0} a_k z^k$.*

Démonstration. Introduisons les séries $R(z) = \sum_k R_k z^k$ et $S(z) = \sum_k S_k z^k$ de telle sorte que $R_k = \sum_{j\geqslant k+1} a_j$ et $S_k = 1 - \sum_{j\leqslant k} b_j$ pour tout k. La série A est strictement croissante sur $[0, R_A]$ avec $R_A > 1$ (comme on l'a vu dans le lemme 1.4) et $A(1) = 1$, si bien que $R(z) = \frac{1-A}{1-z}$ possède seulement une singularité artificielle en $z = 1$ et son rayon de convergence est strictement plus grand que 1 ; celui-ci en fait vaut R_A. Quant à S, telle que $0 \leqslant S_k \leqslant 1$ pour tout k d'après le lemme précédent 1.5, son rayon de convergence est également plus grand que 1.

Introduisons la notation $\odot$ pour le produit de Hadamard des séries, si bien que

$$A \odot B(z) = \sum a_k b_k z^k.$$

La relation (1.2a) se réécrit $2A = F + \alpha A^2 + 2\beta A \odot R + \beta A \odot A$, ou

$$A = \frac{1 - \sqrt{1 - \alpha(F + \beta A \odot (2R + \alpha A))}}{\alpha}.$$

À l'aide de la formule de Cauchy-Hadamard, il est aisé de voir que le produit de Hadamard $X \odot Y$ de deux séries de rayons de convergence respectifs R_X et R_Y possède un rayon de convergence égal à $R_X R_Y$. En effet si $X(z) = \sum_n x_n z^n$ et $Y(z) = \sum_n y_n z^n$, alors

$$R_{X\odot Y} = 1/\left(\varlimsup_{n\to\infty} |x_n y_n|^{\frac{1}{n}}\right) = 1/\left(\varlimsup_{n\to\infty} |x_n|^{\frac{1}{n}}\right) \cdot 1/\left(\varlimsup_{n\to\infty} |y_n|^{\frac{1}{n}}\right) = R_X R_Y.$$

$A \odot (2R + \alpha A)$ possède donc un rayon de convergence strictement plus grand que R_A, et il existe $\varepsilon > 0$ tel que $F + \beta A \odot (2R + \alpha A)$ est une fonction positive strictement croissante sur $[0, R_A + \varepsilon]$, valant $1/\alpha$ en $z = R_A$. En outre, R_A est une racine simple de la fonction $1 - \alpha(F + \beta A \odot (2R + \alpha A))$. La singularité dominante $\rho_A = R_A$ of A correspond au réel pour lequel le radical de A devient nul, autrement dit $A(\rho_A) = \frac{1}{\alpha}$.

La relation (1.2b) se réécrit $B = \beta \dfrac{A \odot S + B \odot R + A \odot B}{1 - \alpha A}$, ou encore

$$B = \beta \frac{A \odot S + B \odot R + A \odot B}{\sqrt{1 - \alpha(F + \beta A \odot (2R + \alpha A))}}. \tag{1.3}$$

On peut déduire de cette expression que B a un rayon de convergence strictement plus grand que 1. En effet, d'une part, la série $1/\sqrt{1 - \alpha(F + \beta A \odot (2R + \alpha A))}$ est analytique de rayon de convergence ρ_A. D'autre part, le numérateur $A \odot S + B \odot R + A \odot B$ a un rayon de convergence strictement plus grand que celui de B : les rayons de convergence de $A \odot S$, $B \odot R$ et $A \odot B$ sont respectivement égaux à $R_A R_S$ (qui vaut au moins $R_A > 1$ puisque $R_S \geqslant 1$), $R_B R_R = R_B R_A$ et $R_A R_B$. Donc si B admettait un rayon de convergence plus petit que 1 (au sens large), alors par une nouvelle application du théorème de Cauchy-Hadamard le membre de droite aurait un rayon de convergence strictement plus grand que celui de B, ce qui est absurde.

Du fait que B possède un rayon de convergence strictement plus grand que 1, on déduit que le numérateur $A \odot S + B \odot R + A \odot B$ possède un rayon de convergence au moins égal à R_A, et qu'il en est de même pour B.

□

Lemme 1.7 $\sum_{k\geqslant 0} b_k = 1$.

Démonstration. Introduisons maintenant les séries $Y_k(z) = \frac{a_k - 2zW_k(z)}{\sqrt{1-4z}}$ qui sont analytiques sur $[0, 1/4[$, et dont la somme $\sum_k Y_k$ vaut $\frac{1-2zT}{\sqrt{1-4z}} = 1$. Les Y_k sont positifs sur $[0, 1/4[$ et vérifient $Y_k(z) = b_k + O(\sqrt{1-4z})$. Pour tout n, nous avons $\sum_{k\leqslant n} Y_k(z) \leqslant 1$, et en faisant tendre z vers $1/4$, nous obtenons $\sum_{k\leqslant n} b_k \leqslant 1$. Comme les b_k sont positifs, la somme $\sum_{k\geqslant 0} b_k$ est convergente, de somme $B(1)$ plus petite que 1 (au sens large).

Posons $x = \sqrt{1-4z}$, si bien que $z = \frac{1-x^2}{4}$. Du développement asymptotique de $W_k(z)$ donné dans le lemme 1.1, on déduit donc l'écriture

$$2zW_k(z) = a_k - b_k x + c_k(x) x^2,$$

où c_k est une fonction bornée de x. En gardant dans (1.1) les termes en x^2, x^3, $\cdots$, x^n, $\cdots$, avec $n \geqslant 2$, nous obtenons

$$\begin{aligned}
&\Bigg(2 - 2\alpha a_0 - 2\beta\Big(1 - \sum_{j=0}^{k} a_j\Big) \\
&\quad +2\alpha b_0 x + 2\beta\Big(1 - \sum_{j=0}^{k} b_j\Big)x - 2\alpha c_0(x)x^2 + 2\beta \sum_{j=0}^{k} c_j(x)x^2 - \beta c_k(x)x^2\Bigg)\, c_k(x) \\
= \;& -f_k + \alpha \sum_{j=0}^{k} b_j b_{k-j} + 2\alpha \sum_{j=1}^{k} a_j c_{k-j}(x) + 2\beta\Big(1 - \sum_{j=0}^{k} b_j\Big)b_k - 2\beta a_k \sum_{j=0}^{k-1} c_j(x) \\
&\quad +\beta b_k^2 - 2\alpha \sum_{j=1}^{k} b_j c_{k-j}(x)x + 2\beta b_k \sum_{j=0}^{k-1} c_j(x)x + \alpha \sum_{j=1}^{k-1} c_j(x)c_{k-j}(x)x^2.
\end{aligned}$$

Introduisons à nouveau les séries $Y_k(z) = \frac{a_k - 2zW_k(z)}{\sqrt{1-4z}}$ déjà définies dans le lemme 1.5. En remarquant que $xc_k(x) = b_k - Y_k(\frac{1-x^2}{4})$, nous obtenons, pour tous entiers i, j tels que $i < j$, l'inégalité

$$\begin{aligned}
\left|\sum_{l=i}^{j} xc_l(x)\right| &\leqslant \sum_{l=i}^{j} x\,|c_l(x)| = \sum_{l=i}^{j} \left|b_l - Y_l\left(\frac{1-x^2}{4}\right)\right| \\
&\leqslant \sum_{l=i}^{j} b_l + Y_l\left(\frac{1-x^2}{4}\right) \leqslant B(1) + 1.
\end{aligned}$$

Le cas $a_0 = 1$ est trivial. Nous pouvons donc supposer $a_0 \neq 1$. Soit ε un réel strictement compris entre 0 et 2β (notons que nous avons aussi $\varepsilon < 2 - 2\alpha a_0$).

Soient K et x_0 tels que, pour $k \geqslant K$ et $0 \leqslant x \leqslant x_0$, on ait à la fois

$$2\beta\big(1 - \sum_{j=0}^{k} a_j\big) < \varepsilon/2$$

et

$$2\alpha b_0 x + 2\beta\big(1 - \sum_{j=0}^{k} b_j\big)x + 2\alpha x|c_0(x)|x + 2\beta\Big|\sum_{j=0}^{k} xc_j(x)\Big|x < \varepsilon/2.$$

Alors pour $k \geqslant K$ et $0 \leqslant x \leqslant x_0$ nous avons

$$\begin{aligned}(2-2\alpha a_0-\varepsilon)\,c_k(x) &\leqslant f_k+\alpha\sum_{j=0}^{k} b_j b_{k-j}+2\alpha\sum_{j=1}^{k} a_j|c_{k-j}(x)| \\ &\quad +2\beta\big(1-\sum_{j=0}^{k} b_j\big)b_k+2\beta a_k\sum_{j=0}^{k-1}|c_j(x)|+\beta b_k^2 \\ &\quad +2\alpha\sum_{j=1}^{k} b_j|c_{k-j}(x)|x+2\beta b_k\sum_{j=0}^{k-1}|c_j(x)|x+\Big|\alpha\sum_{j=1}^{k-1} c_j(x)c_{k-j}(x)x^2\Big| \\ &\leqslant f_k+\alpha\sum_{j=0}^{k} b_j b_{k-j}+2\alpha\sum_{j=1}^{k} a_j|c_{k-j}(x)| \\ &\quad +2\beta\big(1-\sum_{j=0}^{k} b_j\big)b_k+2\beta a_k\sum_{j=0}^{k-1}|c_j(x)|+\beta b_k^2 \\ &\quad +2\alpha\sum_{j=1}^{k} b_j|c_{k-j}(x)|x+2\beta b_k(B(1)+1)+\alpha(B(1)+1)^2.\end{aligned}$$

Pour $x_1 \in]0,x_0[$, soit $D(z)=\sum_{k\geqslant 0} d_k z^k$ la série, dépendant de x_1, définie pour $k<K$ par

$$d_k=\max_{x\in[0,x_1]} c_k(x),$$

et pour $k \geqslant K$ par

$$\begin{aligned}(2-2\alpha a_0-\varepsilon)\,d_k &= f_k+\alpha\sum_{j=0}^{k} b_j b_{k-j}+2\alpha\sum_{j=1}^{k} a_j d_{k-j} \\ &\quad +2\beta\big(1-\sum_{j=0}^{k} b_j\big)b_k+2\beta a_k\sum_{j=0}^{k-1} d_j+\beta b_k^2 \\ &\quad +2\alpha\sum_{j=1}^{k} b_j d_{k-j}x_1+2\beta b_k(B(1)+1)+(B(1)+1)^2.\end{aligned}$$

Il est facile de montrer par récurrence que d_k est plus grand que $|c_k(x)|$ pour tous $k \geqslant 0$ et $x \in [0,x_1]$.

Si on multiplie ces égalités par z^k et qu'on les somme ensuite pour $k \geqslant K$, on obtient l'existence d'un polynôme P_K en z de degré strictement plus petit que K tel que

$$\begin{aligned}(2-2\alpha a_0-\varepsilon)\,D &= P_K+F+\alpha B^2+2\alpha(A-a_0)D+2\beta\frac{1-B}{1-z}\odot B \\ &\quad +2\beta A\odot z\frac{1-D}{1-z}+\beta B\odot B+2\alpha(B-b_0)x_1 D \\ &\quad +2(B(1)+1)B+\alpha(B(1)+1)^2,\end{aligned}$$

ou

$$\begin{aligned}2\big(1-\alpha A-\alpha(B-b_0)x_1-\varepsilon/2\big)D &= P_K+F+\alpha B^2+2\beta\frac{1-B}{1-z}\odot B \\ &\quad +2\beta A\odot z\frac{1-D}{1-z}+\beta B\odot B+2(B(1)+1)B+\alpha(B(1)+1)^2.\end{aligned}$$

Soit R_D le rayon de convergence de D. Le membre de droite possède clairement un rayon de convergence au moins égal à $\min(R_B, R_A \cdot \min(1, R_D))$. Fixons maintenant $x_1 \in]0,1[$ de telle sorte que $x_1 < \frac{1-\alpha-\varepsilon/2}{\alpha((B(1)-b_0)}$. Alors la série $H(z) = 1 - \alpha A(z) - \alpha(B(z) - b_0)x_1 - \varepsilon/2$ satisfait

$$H(1) = 1 - \alpha - \alpha((B(1) - b_0)x_1 - \varepsilon/2 > 0.$$

Comme H est strictement décroissante sur $[0,1]$, il s'ensuit que H est strictement positive sur le même intervalle. En raisonnant par l'absurde et en utilisant la formule de Cauchy-Hadamard, on en déduit que D n'admet pas de singularité dans $[0,1]$ et possède un rayon de convergence strictement plus grand que 1. En conséquence, $\sum_k c_k(x)$ converge normalement et donc uniformément sur $[0, x_1]$, et nous pouvons conclure que, sur $[0, x_1]$, $\sum_k W_k(\frac{1-x^2}{4}) = \sum_k a_k - \sum_k b_k x = 1 - x$ et donc que $\sum_k b_k = 1$. □

Nous sommes maintenant en mesure d'établir la nature de la loi limite de X_n dans le cas général. Soit $\mathcal{T}((+,\alpha),(\min,\beta),I)$ l'ensemble pondéré des arbres de Catalan tels que, d'une part, les opérateurs sont égaux à "+" avec probabilité α et à "min" avec probabilité $\beta = 1 - \alpha$, et d'autre part, les opérandes sont choisis au hasard dans un ensemble fini pondéré d'entiers I.

Cela nous permet d'établir la preuve du théorème annoncé en début de paragraphe.

Preuve du théorème 1.1 Pour tout k on a $W_k = 2a_k - 2b_k\sqrt{1-4z} + O(1-4z)$, avec a_k et b_k tous les deux strictement positifs ou égaux à zéro — le dernier cas correspond à $W_k = 0$. On admettra que W_k ne possède pas de singularité dominante autre que $z = \frac{1}{4}$. Alors par analyse de singularité [82], nous en déduisons que pour tout k,

$$[z^n]W_k = \frac{b_k}{\sqrt{\pi n^3}}4^n + O\left(\frac{1}{n^{5/2}}\right) \text{ et } \lim_{n\to+\infty} \frac{[z^n]W_k}{C_n} = b_k,$$

où $\sum_{k\geqslant 0} b_k = 1$. Donc la valuation a une limite discrète X définie par les probabilités

$$\mathbf{Pr}[X = k] = b_k$$

pour tout k.

Les suites (a_k) et (b_k) sont définies par leurs premiers termes calculés dans la preuve de la proposition 1.1, et par les récurrences indiquées dans la proposition 1.2. □

À défaut de trouver une expression générale pour tous les b_k, nous pouvons établir la nature asymptotique de la suite (b_k), comme décrite dans la proposition 1.2.

Preuve de la proposition 1.2 De $\sum_k b_k = 1$, et de ce que B est une série analytique au-delà de 1 avec $B(1) = 1$, on déduit que 1 est racine au moins simple de $1 - B$. Il s'ensuit que la série $S(z) = \frac{1-B}{1-z}$ introduite dans la preuve du lemme 1.6 admet seulement une singularité artificielle en 1 et que son rayon de convergence vaut en fait R_B. Il s'ensuit que $A \odot S$ admet un rayon de convergence strictement plus grand que celui de A et que le numérateur $A \odot S + B \odot R + A \odot B$ dans l'expression (1.3) de B est analytique au-delà de R_A.

On en déduit que la singularité dominante ρ_B est un zéro simple du radical du numérateur, et nous obtenons l'asymptotique de B comme nous l'avions affirmé par analyse de singularité. □

Dans la figure 1.1, on peut voir comment les grandeurs ρ_B et C, définies dans la proposition précédente, varient avec α pour $\mathbf{Pr}[X_0 = 0] = \mathbf{Pr}[X_0 = 1] = 1/2$. Notons que, heuristiquement, $\lim_{\alpha\to 0} \rho_B = +\infty$ et $\lim_{\alpha\to 1} \rho_B = 1$, comme on pouvait s'y attendre étant donné les lois limites pour les cas extrêmes, à savoir une distribution de masse de Dirac pour $\alpha = 0$, et une distribution gaussienne pour $\alpha = 1$.

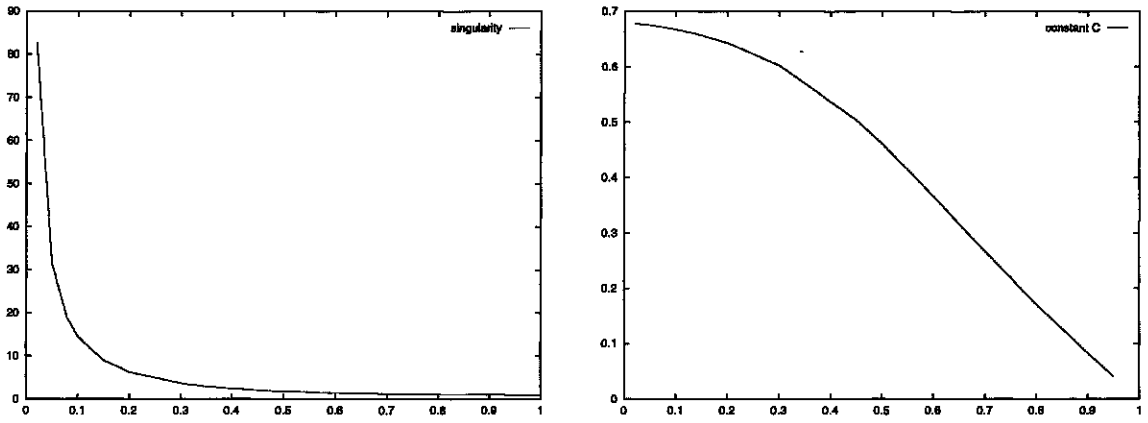


FIG. 1.1 – ρ_B en fonction de α et C en fonction de α.

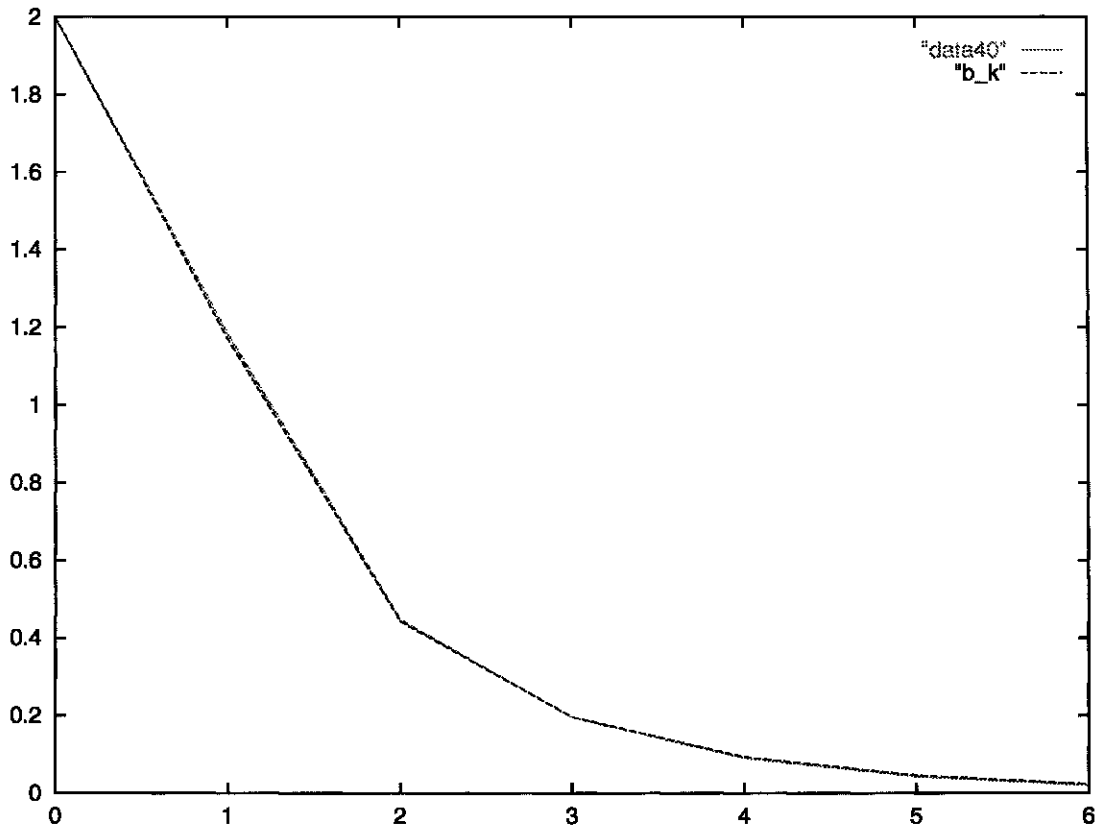


FIG. 1.2 – Superposition de b_k et de la distribution de la valuation pour $n = 40$ dans le cas min et + équiprobables et des opérandes 0 et 1 équiprobables.

1.1.4 Interprétation probabiliste des a_k

On a vu que la somme des a_k valait 1, avec $a_k \geqslant 0$ pour tout k, si bien que $A(x)$ est une série génératrice de probabilité. En raisonnant comme dans [31], on peut trouver une interprétation probabiliste de cette distribution.

Lemme 1.8 *La probabilité d'obtenir un arbre de valuation k en partant d'un processus de branchement simple critique, avec probabilité $\frac{1}{2}$ d'obtenir zéro ou deux fils, et un étiquetage des nœuds internes et externes suivant le modèle étudié plus haut, est donnée par la suite (a_k) définie comme dans (1.1).*

Démonstration. Soit τ un arbre de Galton-Watson correctement étiqueté, *i.e.* les nœuds internes comportent un "+" avec probabilité α, un "min" avec probabilité $1-\alpha$, et chaque nœud externe comporte un k avec probabilité f_k. Soit $\widetilde{\tau}$ l'arbre non étiqueté de même structure que τ. Soit $|\tau|_t$ le nombre total de nœuds de l'arbre τ, et $|\tau|_i$ et $|\tau|_e$ les nombres de nœuds respectivement internes et externes. (Rappelons les relations $|\tau|_t = 2|\tau|_i + 1$ et $|\tau|_e = |\tau|_i + 1$).

Dans ce processus critique de Galton-Watson, la probabilité d'obtenir un arbre $\widetilde{\tau}$ de taille $|\tau|_t$ est donnée par

$$\mathbf{Pr}(\widetilde{\tau}) = \frac{1}{2^{|\tau|_t}}.$$

La probabilité d'un arbre τ de structure $\widetilde{\tau}$, d'étiquettes internes o_i et externes e_i, est donnée par $\mathbf{Pr}(\tau) = \prod_{i=1}^{|\tau|_e} p_{e_i} \prod_{i=1}^{|\tau|_i} p_{o_i} \mathbf{Pr}(\widetilde{\tau})$. Maintenant, la probabilité $\pi(k)$ qu'un arbre de Galton-Watson étiqueté soit de valuation k est donnée par :

$$\begin{aligned}
\pi(k) &= \sum_{\tau,\mathrm{val}(\tau)=k} \mathbf{Pr}(\tau) \\
&= \sum_{\tau,\mathrm{val}(\tau)=k} \prod_{i=1}^{|\tau|_e} p_{e_i} \prod_{i=1}^{|\tau|_i} p_{o_i} \mathbf{Pr}(\widetilde{\tau}) \\
&= \sum_{\tau,\mathrm{val}(\tau)=k} \sum_{n\geqslant 0} \sum_{\tau,|\tau|_i=n} \prod_{i=1}^{|\tau|_e} p_{e_i} \prod_{i=1}^{|\tau|_i} p_{o_i} \frac{1}{2^{2n+1}} \\
&= \frac{1}{2} \sum_{\tau,\mathrm{val}(\tau)=k} \sum_{n\geqslant 0} \sum_{\tau,|\tau|_i=n} \prod_{i=1}^{|\tau|_e} p_{e_i} \prod_{i=1}^{|\tau|_i} p_{o_i} \left(\frac{1}{4}\right)^n \\
&= \frac{1}{2}[u^k]W(1/4,u) \\
&= a_k.
\end{aligned}$$

□

1.1.5 Un mot sur les arbres max, +

Notons que le comportement de la valuation d'arbres arithmétiques avec des max et des "+" est complètement différent. En effet, si W_k désigne la série génératrice des arbres de valuation k, alors les W_k n'ont pas la même singularité. Plus précisément, si on note ρ_k la singularité de W_k, alors $\rho_k > \rho_{k'}$ pour $k < k'$, si bien qu'on n'a plus de limite discrète.

Pour l'instant, même la moyenne de la valuation reste encore inconnue. Nous savons seulement qu'elle tend vers l'infini avec la taille de l'arbre, puisque la propriété ci-dessus signifie que

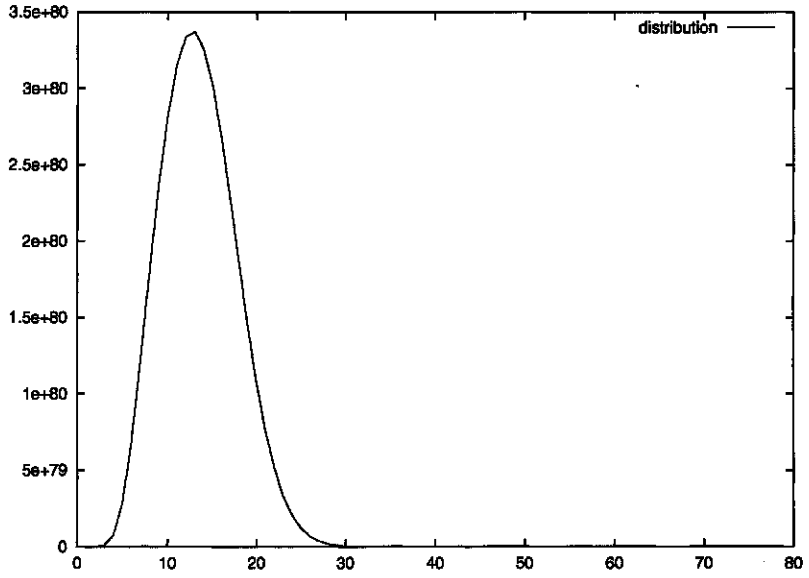


FIG. 1.3 – Distribution de la valuation max, + pour des arbres de taille 70.

le nombre d'arbres de valuations k est asymptotiquement exponentiellement petit par rapport au nombre d'arbres de valuations k' si $k < k'$.

Des simulations montrent une loi asymétrique, si bien que la limite pourrait ne pas être gaussienne, si cette asymétrie est conservée à la limite. La figure 1.3 montre la distribution de la valuation d'arbres avec opérateurs équiprobables max et "+", et opérandes équiprobables 0 et 1 (les arbres sont comptés sans pondération).

La différence de comportement entre les jeux d'opérateurs $\{\min, +\}$ et $\{\max, +\}$ peut *a priori* surprendre, puisque ces jeux d'opérateurs mènent habituellement aux mêmes types de résultats [4]. Néanmoins, dans un arbre de Catalan, des sous-arbres de taille finie, même avec un arbre de taille arbitrairement grande, apparaissent avec des probabilités non nulles, (la taille des sous-arbres suit une loi limite discrète), si bien que l'opérateur min écrase les valuations, ce qui n'est pas le cas de l'opérateur max.

1.2 Le paramètre 2-bud

La méthodologie permet de trouver la distribution limite d'une valuation existant déjà dans la littérature, à savoir la valuation 2-bud introduite par Vincent Blondel dans [16] : on y trouve des considérations algébriques sur ce paramètre, mais pas d'étude statistique. Ce paramètre possède plusieurs interprétations, mais nous les omettons ici par manque de place, car elles nécessitent des définitions un peu compliquées.

Ce paramètre ressemble beaucoup au nombre de Strahler, dont la définition diffère de 2-bud en remplaçant min par max.

Définition 1.1 *Le paramètre 2-bud d'un arbre de Catalan $T = \circ(T_1, T_2)$ est une valuation v définie par $v(T) = 1$ si $T = \square$, et $v(T) = [v(T_1) = v(T_2)] + \min(v(T_1), v(T_2))$ si $T = \circ(T_1, T_2)$.*

Autrement dit, $v(T)$ vaut $1 + v(T_1)$ si $v(T_1) = v(T_2)$, et $\min(v(T_1), v(T_2))$ sinon.

Théorème 1.3 *Soit, pour tout $k \in \mathbb{N}^*$, la série génératrice W_k des arbres de valuation k. Alors pour tout k, pour z tendant vers $\frac{1}{4}$, on a le développement asymptotique :*

$$W_k = 2a_k - 2b_k\sqrt{1-4z} + O(1-4z),$$

où $(a_k)_{k\in\mathbb{N}^*}$ et $(b_k)_{k\in\mathbb{N}^*}$ vérifient $\sum_k a_k = \sum_k b_k = 1$.

Démonstration. Posons

$$W(z) = \sum_k W_k(z) = \frac{1-\sqrt{1-4z}}{2z}$$

la série génératrice des arbres binaires de Catalan. On a

$$\begin{aligned} W_1 &= 1 + 2zW_1W_{>1} = 1 + 2zW_1(W - W_1) \\ &= 1 + 2zWW_1 - 2zW_1^2 \\ &= \frac{1 - 2zW - \sqrt{(1-2zW)^2 + 8z}}{-4z}, \end{aligned}$$

et généralement pour tout k,

$$\begin{aligned} W_k &= zW_{k-1}^2 + 2zW_kW_{>k} \\ &= zW_{k-1}^2 + 2zW_k(W - W_{\leqslant k-1} - W_k) \\ &= zW_{k-1}^2 + 2z(W - W_{\leqslant k-1})W_k - 2zW_k^2 \\ &= \frac{-(1-2z(W-W_{\leqslant k-1})) + \sqrt{(1-2z(W-W_{\leqslant k-1}))^2 + 8z^2W_{k-1}^2}}{4z}. \end{aligned} \tag{1.4}$$

Par construction, pour tout k les coefficients de W_k en tant que série en z sont majorés par ceux de W, si bien que W_k possède un rayon de convergence au moins égal à $\frac{1}{4}$, en laquelle il admet au pire une singularité de type quadratique.

Écrivons $W_k = 2a_k - 2b_k\sqrt{1-4z} + O(1-4z)$, où b_k peut *a priori* être nul. Par identification en le terme constant et le terme en $\sqrt{1-4z}$ dans le développement asymptotique de (1.4) autour de $z = 1/4$, on obtient :

$$\begin{cases} 2a_k &= a_{k-1}^2 + 2a_k(1 - \sum_{j=0}^{k-1} a_j) - 2a_k^2, \\ -2b_k &= -2a_{k-1}b_{k-1} + 2(-b_k(1 - \sum_{j=0}^{k-1} a_j) - a_k(1 - \sum_{j=0}^{k-1} b_j)) + 4a_kb_k, \end{cases}$$

soit :

$$\begin{cases} a_k^2 + \sum_{j=0}^{k-1} a_ja_k - \frac{1}{2}a_{k-1}^2 = 0, \\ a_kb_k + \sum_{j=0}^{k} a_jb_k - a_{k-1}b_{k-1} - a_k(1 - \sum_{j=0}^{k-1} b_j) = 0, \end{cases},$$

et finalement :

$$\begin{cases} 2a_k &= -\sum_{j=0}^{k-1} a_j + \sqrt{\left(\sum_{j=0}^{k-1} a_j\right)^2 + 2a_{k-1}^2}, \\ \left(a_k + \sum_{j=0}^{k} a_j\right) b_k &= a_{k-1}b_{k-1} + (1 - \sum_{j=0}^{k-1} b_j)a_k. \end{cases}$$

Notons qu'on a $a_1 = \sqrt{2}/2$, $b_1 = 1/2$.

Montrons que la série entière $\sum a_kz^k$ possède un rayon de convergence infini. On a :

$$\begin{cases} 2a_k &= a_{k-1}^2 + 2a_k(1 - \sum_{j=0}^{k-1} a_j) - 2a_k^2 \\ &= \sum_{j=0}^{k-1} a_j \left(-1 + \sqrt{1 + 2a_{k-1}^2}\right), \end{cases}$$

d'où, en utilisant $\sqrt{1+y} \leqslant 1 + y/2$ pout tout y :

$$a_k \leqslant \frac{1}{2} \sum_{j=0}^{k-1} a_j a_{k-1}^2 \leqslant \frac{a_{k-1}^2}{2}.$$

Or, puisque $\sum_k a_k = \sum_k W_k(1/4)$, on a $\lim_{k\to\infty} a_k = 0$ et, pour tout $\varepsilon > 0$, il existe k_0 tel que, pour $k \geqslant k_0$, $a_{k-1} \leqslant \varepsilon$, et par suite $a_k \leqslant (\varepsilon/2)^{k-k_0} a_{k_0}$. Autrement dit $\overline{\lim}_k a_k \varepsilon^{-k}$ est fini pour tout $\varepsilon > 0$, donc $\sum a_k z^k$ a un rayon de convergence infini.

Montrons $\sum_k b_k \leqslant 1$. Posons pour cela $Y_k(z) = \frac{a_k - 2zW_k}{\sqrt{1-4z}}$, si bien que $Y_k(\frac{1-x^2}{4}) = b_k - c_k(x)x$. Nous avons :

$$\begin{aligned} \sum_k Y_k(z) &= \sum_k \frac{a_k - 2zW_k}{\sqrt{1-4z}} \\ &= \frac{1 - 2zW}{\sqrt{1-4z}} = 1, \end{aligned}$$

avec $Y_k(z) > 0$ sur $[0, 1/4]$ car, d'une part, $Y(1/4) = b_k$ et d'autre part, pour $z < 1/4$, on a $\sqrt{1-4z} > 0$ et $2zW_k < a_k$. On en déduit pour tous n, z :

$$\sum_{k\leqslant n} Y_k(z) \leqslant 1,$$

d'où en faisant tendre z vers $1/4$:

$$\sum_{k\leqslant n} Y_k(1/4) \leqslant 1,$$

soit

$$\sum_{k\leqslant n} b_k \leqslant 1.$$

On en déduit que la série de terme général b_k est convergente vers une somme $B(1) \leqslant 1$.

Montrons que le rayon de convergence de la série B est infini. Soient $\varepsilon \in]0,1[$ et k_0 entier tels que $\sum_{j\leqslant k} a_k > 1 - \varepsilon$ pour tout $k \geqslant k_0$. De

$$\sum_{j=0}^{k} a_j b_k = a_{k-1} b_{k-1} + a_k \left(1 - \sum_{j=0}^{k} b_j\right),$$

sachant qu'on a $0 \leqslant b_{k-1} \leqslant 1$ et $0 \leqslant 1 - \sum_{j=0}^{k} b_j \leqslant 1$, on tire pour $k \geqslant k_0$:

$$b_k \leqslant \frac{1}{1-\varepsilon} (a_{k-1} + a_k).$$

On en déduit que le rayon de convergence de B est plus grand que celui de A et est donc aussi infini.

Posons maintenant pour tout k, par commodité de calcul, $Z_k = 2zW_k$ et $x = \sqrt{1-4z}$, si bien que $z = \frac{1-x^2}{4}$. Du développement asymptotique de W_k autour de $z = \frac{1}{4}$, on déduit l'existence d'une fonction c_k de x, bornée au voisinage de $x = 0$, telle que

$$Z_k\left(\frac{1-x^2}{4}\right) = a_k - b_k x + c_k(x) x^2.$$

Notons $Z = \sum_k Z_k = 1 - \sqrt{1-4z}$ et $Z_{\leqslant k} = \sum_{j \leqslant k} Z_j$. Nous avons :

$$Z_k = \frac{1}{2} Z_{k-1}^2 + (Z - Z_{\leqslant k}) Z_k - Z_k^2,$$

soit :

$$\begin{aligned} a_k - b_k x + c_k(x) x^2 &= \frac{1}{2}\big(a_{k-1} - b_{k-1}x + c_{k-1}(x)x^2\big)\big(a_{k-1} - b_{k-1}x + c_{k-1}(x)x^2\big) \\ &\quad + \Big(\big(1 - \sum_{j=0}^{k-1} a_j\big) - \big(1 - \sum_{j=0}^{k-1} b_j\big)x - \sum_{j=0}^{k-1} c_j(x)x^2 \Big) \cdot \big(a_k - b_k x + c_k(x)x^2\big) \\ &\quad - \big(a_k - b_k x + c_k(x)x^2\big)\big(a_k - b_k x + c_k(x)x^2\big). \end{aligned} \tag{1.5}$$

Montrons $B(1) = 1$. Pour cela on va montrer la convergence uniforme de la somme des $c_k(x)$ au voisinage de $x = 0$. Remarquons tout d'abord pour tout $x \in [0,1]$ et tout sous-ensemble E de $\mathbb{N}$:

$$x \left| \sum_{k \in E} c_k(x) \right| = \left| \sum_{k \in E} b_k - Y_k \right| \leqslant \sum_{k \in E} b_k + \sum_{k \in E} Y_k \leqslant B(1) + 1 \tag{1.6}$$

En éliminant les termes en x^0 et en x de l'égalité (1.5) et en divisant ce qui reste par x^2 on obtient

$$\begin{aligned} c_k(x) &= \frac{1}{2}(2a_{k-1}c_{k-1}(x) + b_{k-1}^2 - 2xb_{k-1}c_{k-1}(x) + c_{k-1}^2(x)x^2) \\ &\quad + \Big(1 - \sum_{j=0}^{k-1} a_j\Big)c_k(x) - \Big(1 - \sum_{j=0}^{k-1} b_j\Big)b_k - a_k \sum_{j=0}^{k-1} c_j(x) \\ &\quad + x c_k(x) \sum_{j=0}^{k-1} b_j + x b_k \sum_{j=0}^{k-1} c_j - x^2 c_k(x) \sum_{j=0}^{k-1} c_j - 2a_k c_k(x) - b_k^2 + 2x b_k c_k(x) - c_k(x)^2 x^2 \end{aligned}$$

d'où :

$$\begin{aligned} &\Big(\sum_{j=0}^{k} a_k + a_k x\Big(1 - \sum_{j=0}^{k} b_j - b_k\Big) + x^2 \sum_{j=0}^{k} c_j(x)\Big)c_k(x) \\ = \; & a_{k-1}c_{k-1}(x) + \frac{1}{2}b_{k-1}^2 - x b_{k-1}c_{k-1}(x) + \frac{1}{2}c_{k-1}^2(x)x^2 \\ & - \Big(1 - \sum_{j=0}^{k-1} b_j\Big)b_k - a_k \sum_{j=0}^{k-1} c_j(x) + x b_k \sum_{j=0}^{k-1} c_j(x) - b_k^2. \end{aligned}$$

Soient $\varepsilon \in]0,1[$, $x_0 \in]0,1[$ et $K \in \mathbb{N}$ tels que, pour $x \leqslant x_0$ et $k \geqslant K$ on ait

$$\sum_{j=0}^{k} a_k + a_k x\Big(1 - \sum_{j=0}^{k} b_j - b_k\Big) + x^2 \sum_{j=0}^{k} c_j(x) > 1 - \varepsilon.$$

Pour $k \geqslant K$ et $0 \leqslant x \leqslant x_0$, on a en utilisant la majoration 1.6 :

$$\begin{aligned}(1-\varepsilon)|c_k(x)| &\leqslant a_{k-1}|c_{k-1}(x)| + \frac{1}{2}b_{k-1}x^2 + xb_{k-1}|c_{k-1}(x)| + \frac{1}{2}c_{k-1}^2(x)x^2 \\ &\quad + \Big(1 - \sum_{j=0}^{k-1} b_j\Big)b_k + a_k \sum_{j=0}^{k-1} |c_j(x)| + xb_k \Big|\sum_{j=0}^{k-1} c_j(x)\Big| + b_k^2. \\ &\leqslant a_{k-1}|c_{k-1}(x)| + \Big(\frac{x_0^2}{2} + 2B(1) + 2\Big)b_{k-1} + a_k \sum_{j=0}^{k-1} |c_j(x)| + 2b_k + \frac{B(1)+1}{2}x|c_k|.\end{aligned}$$

Définissons maintenant la série $D(z) = \sum_k d_k z^k$ par :

$$d_k = \begin{cases} \max\limits_{x \leqslant x_0} |c_k(x)| & \text{si } k < K, \\ \dfrac{1}{1-\varepsilon}\Big(a_{k-1}d_{k-1} + \Big(\dfrac{x_0^2}{2} + 2B(1) + 2\Big)b_{k-1} \\ \quad + a_k \sum_{j=0}^{k-1} d_j + 2b_k + \frac{1}{2}(B(1)+1)d_{k-1}x_0\Big) & \text{si } k \geqslant K \end{cases}$$

On a $|c_k(x)| \leqslant d_k$ pour $x \leqslant x_0$ et $k < K$, et par récurrence on montre que cette inégalité reste vérifiée pour $k \geqslant K$.

Par définition des d_k pour $k \geqslant K$, il existe un polynôme $P_K(z)$ de degré au plus $K-1$ tel que :

$$D = \frac{1}{1-\varepsilon}\left(z(A \odot D) + z\Big(\frac{x_0^2}{2} + 2B(1) + 2\Big)B + A \odot \Big(z\frac{1-D}{1-z}\Big) + 2B + \frac{B(1)+1}{2}x_0 zD\right) + P_K,$$

soit :

$$\begin{aligned}&\Big(1 - \frac{(B(1)+1)x_0}{2(1-\varepsilon)}z\Big)D \\ &= \frac{1}{1-\varepsilon}\left(z(A \odot D) + z\Big(\frac{x_0^2}{2} + 2B(1) + 2\Big)B + A \odot \Big(z\frac{1-D}{1-z}\Big) + 2B\right) + P_K.\end{aligned}$$

Il existe des constantes μ et ν telles que $d_k < \mu d_{k-1} + \nu$ pour tout $k \geqslant K$, donc D a un rayon de convergence non nul. On en déduit, à l'aide de la formule de Cauchy-Hadamard, que $A \odot D$ et $A \odot \left(z\frac{1-D}{1-z}\right)$ ont un rayon de convergence infini. Rappelons que B a aussi un rayon de convergence infini. Il s'ensuit que le membre de droite est une série entière de rayon de convergence infini. Il en est donc de même du membre de gauche, et le rayon de convergence R_D de D vérifie :

$$R_D \geqslant \frac{2(1-\varepsilon)}{(B(1)+1)x_0}.$$

Pour x_0 suffisamment petit, on a $R_D > 1$.

Finalement, la somme $\sum_k c_k(x)$ converge uniformément sur $[0, x_0]$ pour un tel x_0. Il en est de même pour la somme $\sum_k Y_k(\frac{1-x^2}{4})$ et en particulier $\sum_k Y_k(1/4) = \lim_{x \to 0+}(\sum_k Y_k(\frac{1-x^2}{4})) = 1$, *i.e.* $\sum b_k = 1$. □

Lemme 1.9 *Tous les arbres de taille paire ont un paramètre 2-bud égal à 1.*

Démonstration. De l'expression trouvée précédemment pour W_1, série génératrice des arbres de valuation 1, on déduit facilement

$$\begin{aligned} W_1(z) &= \frac{\sqrt{1+4z}-\sqrt{1-4z}}{4z} \\ &= \frac{W(z)+W(-z)}{2}, \end{aligned}$$

ce qui est la partie paire de la série génératrice complète des arbres de Catalan binaires. On en déduit que tous les arbres de taille paire ont un paramètre 2-bud égal à 1. □

Théorème 1.4 *Lorsqu'on fait tendre n vers l'infini par valeurs impaires, la valuation 2-bud des arbres de taille n tend vers une limite discrète.*

Démonstration. Montrons par récurrence sur k que W_k est une fonction impaire pour $k \geqslant 2$, au sens où $W_k(z) = -W_k(z)$. Pour $k = 2$ on a facilement

$$W_2(z) = -\frac{\sqrt{1-4z}+\sqrt{1+4z}-\sqrt{6-2\sqrt{1-4z}\sqrt{1+4z}}}{8z},$$

ce qui montre la propriété voulue au rang 2. Soit maintenant $k \geqslant 3$ et supposons la propriété prouvée jusqu'au rang $k-1$ avec $k \geqslant 3$. D'après le lemme 1.9, W_1 est la partie paire de W donc $W - W_1$ est impaire. On en déduit, en utilisant en outre la propriété de récurrence, que la série

$$W - W_{\leqslant k-1} = (W - W_1) - \sum_{j=2}^{k-1} W_j$$

est une somme de séries impaires et est donc impaire. Les séries $z\,(W - W_{\geqslant k-1})$ et $z^2 W_{k-1}^2$ sont donc paire et l'expression

$$W_k = \frac{-(1-2z(W-W_{\leqslant k-1}))+\sqrt{(1-2z(W-W_{\leqslant k-1}))^2+8z^2W_{k-1}^2}}{4z} \tag{1.7}$$

montre que W_k est impaire.

Montrons maintenant que pour tout k les seules singularités de W_k sur le cercle $|z| = \frac{1}{4}$ sont $-\frac{1}{4}$ et $\frac{1}{4}$. On sait déjà que c'est vrai pour $k = 1$ puisqu'alors $W_1(z) = \frac{1}{2}\left(W(z)+W(-z)\right)$. Pour montrer que W_k n'admet pas de singularité autre que $-\frac{1}{4}$ et $\frac{1}{4}$ pour $|z| = \frac{1}{4}$, il suffit d'après l'expression (1.7) que le radical $(1-2z(W-W_{\leqslant k-1}))^2 + 8z^2W_{k-1}^2$ ne s'annule pas sur $|z| = \frac{1}{4}$. Or pour $|z| = \frac{1}{4}$ (et même pour $|z| \leqslant \frac{1}{4}$) on a d'une part

$$\begin{aligned} |W_{k-1}^2(z)| &\leq W_{k-1}^2(|z|) \\ &\leq a_{k-1}^2 \\ \left|8z^2W_{k-1}^2\right| &\leq \frac{1}{2}a_{k-1}^2 < \frac{1}{4} \text{ puisque } a_{k-1} \leqslant 1 - a_1 = 1 - \frac{\sqrt{2}}{2} < \frac{1}{2}, \end{aligned}$$

et d'autre part

$$\begin{aligned}
|1-2z(W-W_{\leqslant k-1})| &\geq 1-2|z|\cdot|W-W_{\leqslant k-1}| \\
&\geq 1-2\frac{1}{4}\cdot 2\sum_{j\geq k} a_j \\
&\geq 1-\left(1-\sum_{j\leq k-1} a_j\right) \\
&\geq \sum_{j\leq k-1} a_j \geq a_1 = \frac{\sqrt{2}}{2} \\
|1-2z(W-W_{\leqslant k-1})|^2 &\geq \frac{1}{2}.
\end{aligned}$$

Il est donc impossible d'avoir pour $|z| \leq \frac{1}{4}$

$$(1-2z(W-W_{\leqslant k-1}))^2 = -8z^2W_{k-1}^2,$$

puisque les membres de part et d'autre de cette égalité diffèrent en module d'au moins $\frac{1}{4}$, ce qui montre que le radical $(1-2z(W-W_{\leqslant k-1}))^2+8z^2W_{k-1}^2$ ne peut pas s'annuler pour $|z| \leq \frac{1}{4}$.

Les séries W_k pour $k \geqslant 2$ sont impaires et vérifient donc, en leurs singularités dominantes $\frac{1}{4}$ et $-\frac{1}{4}$,

$$W_k(z) \underset{z\to\frac{1}{4}}{=} 2a_k - 2b_k\sqrt{1-4z} + O(1-4z),$$

et

$$W_k(z) \underset{z\to-\frac{1}{4}}{=} -2a_k + 2b_k\sqrt{1+4z} + O(1+4z).$$

Du point de vue asymptotique et de l'analyse de singularité, les contributions des singularités $\frac{1}{4}$ et $-\frac{1}{4}$ s'additionnent donc et

$$[z^{2n+1}]W_k(z) \sim 2b_k\frac{4^{2n+1}}{\sqrt{\pi(2n+1)^3}}.$$

Il s'ensuit, connaissant l'asymptotique des nombres de Catalan $[z^{2n+1}]W(z) = C_{2n+1}$,

$$\frac{[z^{2n+1}]W_k(z)}{[z^{2n+1}]W(z)} = 2b_k.$$

On a vu précédemment que d'une part $\sum_{k\geqslant 1} b_k = 1$ et d'autre part $b_1 = \frac{1}{2}$, donc $\sum_{k\geqslant 2} 2b_k = 1$. On peut donc conclure que, pour n tendant vers l'infini par valeurs impaires, la valuation 2-bud des arbres de taille n admet une distribution limite discrète donnée par $(2b_k)_{k\geqslant 2}$. □

Il semble encore difficile de trouver la nature asymptotique exacte de la suite (b_k), mais on peut facilement montrer que la suite décroît plus qu'exponentiellement. Cette décroissance rapide est confirmée par les valeurs numériques des premiers b_k : b_5 est de l'ordre de 10^{-5} et b_6 de l'ordre de 10^{-12}.

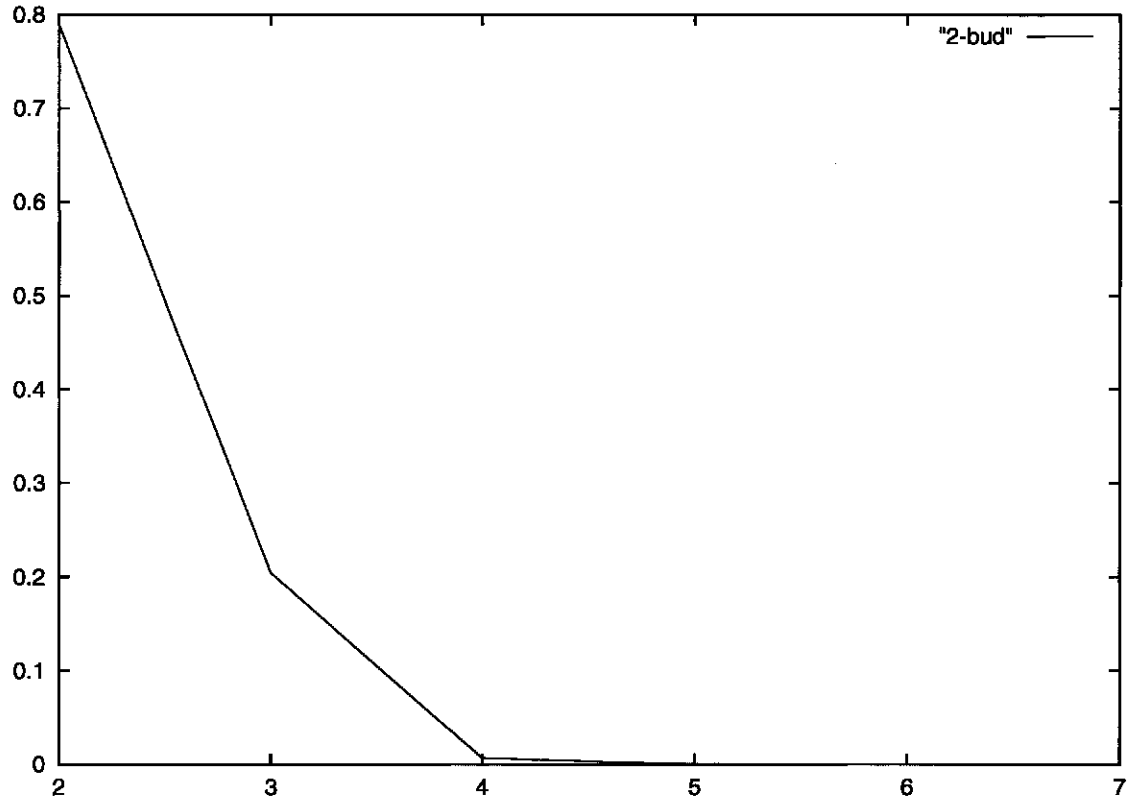


FIG. 1.4 – Distribution limite du paramètre 2-bud.

1.3 Arbres avec opérateurs "+" et "−"

Soit I l'ensemble fini d'entiers qu'on peut placer aux nœuds externes. Chaque entier $k \in I$ apparaît avec probabilité f_k à chaque nœud externe.

Dans cette section, la valuation est définie par

$$v(t) = \begin{cases} k & \text{si } t = \boxed{k} \text{ pour } k \in I, \\ v(t_1) - v(t_2) & \text{si } t = \begin{array}{ccc} & - & \\ / & & \backslash \\ t_1 & & t_2 \end{array}, \\ v(t_1) + v(t_2) & \text{si } t = \begin{array}{ccc} & + & \\ / & & \backslash \\ t_1 & & t_2 \end{array}. \end{cases}$$

Nous avons déjà vu dans la section 1.1 que les arbres munis du seul opérateur "+" ont une valuation de distribution limite gaussienne. Nous allons voir que les arbres munis du seul opérateur "−" aboutissent aussi à une limite gaussienne, et qu'un mélange des deux opérateurs aboutit toujours à une limite gaussienne.

1.3.1 Arbres avec opérateurs "−" et opérandes dans l'ensemble fini I

Bien que la loi limite puisse être déterminée par une méthode plus générale utilisée ci-après, mettant en jeu un système d'équations, nous traitons le cas particulier avec des opérateurs "−" parce que d'une part, il peut être traité avec des méthodes élémentaires, et d'autre part, parce que ce cas introduit notre méthodologie. Nous reprendrons en outre certaines expressions de séries génératrices dans le chapitre 6, où nous étudierons une valuation définie de la même manière récursivement, mais en mettant cette fois des lois sans variance aux feuilles. Dans le chapitre 3, nous démontrerons à nouveau le caractère gaussien de la limite de la distribution de Narayana définie ci-dessous, en utilisant des martingales.

Loi limite

Nous montrons le caractère gaussien de notre variable aléatoire à l'aide d'une analyse de singularité bivariée et du théorème des quasi-puissances de Hwang.

Proposition 1.2 *Si v est définie par*

$$v(t) = \begin{cases} k & \textit{si } t = \boxed{k} \textit{ avec } k \in I, \\ v(t_1) - v(t_2) & \textit{si } t = \begin{array}{ccc} & - & \\ / & & \backslash \\ t_1 & & t_2 \end{array}, \end{cases}$$

alors $\dfrac{X_n}{\sqrt{n\left(\mathbf{Var}(F) + \frac{1}{2}F'(1)^2\right)}}$ *converge en distribution vers* $\mathcal{N}(0,1)$. *(L'espérance* $\mathbf{E}[X_n]$ *est asymptotiquement négligeable devant l'écart-type de* X_n*).*

Démonstration. Pour tout arbre t de sous-arbres t_1 et t_2, sa valuation est $v(t) = v(t_1) - v(t_2)$ et l'arbre t' de sous-arbres t_2 et t_1 a pour valuation $v(t') = -v(t)$. C'est pourquoi on introduit la série bivariée $H(z,u) = W(z,u) - F(u)$ des arbres de taille non nulle énumérés par taille et valuation, qui vérifie donc $H(z,u) = H(z,u^{-1})$. En notant $\overline{F}(u) = F(1/u)$, on déduit donc de

$$H(z,u) = z(F(u) + H(z,u))(F(u^{-1}) + H(z,u^{-1}))$$

une expression pour H :

$$\begin{aligned} H &= z(F+H)(\overline{F}+H) = zH^2 + z(F+\overline{F})H + zF\overline{F} \\ &= \frac{1 - z(F+\overline{F}) - \sqrt{\left(1 - z(F+\overline{F})\right)^2 - 4z^2F\overline{F}}}{2z} \\ H &= \frac{1 - z(F+\overline{F})}{2z} \\ &\quad - \frac{\sqrt{\left(1 - z(F+\overline{F}) + 2z\sqrt{F\overline{F}}\right)\left(1 - z(F+\overline{F}) - 2z\sqrt{F\overline{F}}\right)}}{2z}, \end{aligned}$$

si bien qu'on peut écrire zH sous la forme

$$zH = A(z) + B(z)\sqrt{C(z)},$$

avec

$$A(z) = \frac{1 - z(F+\overline{F})}{2},$$

$$B(z) = -\frac{1}{2}\sqrt{\left(1 - z(\sqrt{F} - \sqrt{\overline{F}})^2\right)},$$

et

$$C(z) = \left(1 - z(\sqrt{F} + \sqrt{\overline{F}})^2\right).$$

La série H, considérée comme fonction de z, admet pour singularité dominante $\rho(u) = (\sqrt{F} + \sqrt{\overline{F}})^{-2}$, avec ρ analytique dans un voisinage de $u = 1$, où il vaut $1/4$. En utilisant le schéma méromorphe du théorème des quasi-puissances de Hwang (p.56 de [85]), rappelé en annexe A, et le fait que $\rho'(1) = 0$, comme on peut le vérifier facilement à partir de l'expression de $\rho(u)$, nous obtenons que la valuation des arbres considérés est asymptotiquement gaussienne, avec $\mathbf{E}[X_n] = -n\rho'(1)/\rho(1) + O(1) = O(1)$ et $\mathbf{Var}[X_n] = -n\rho''(1)/\rho(1) + O(1)$.

Un calcul élémentaire permet de montrer que

$$\mathbf{Var}\left(\left(\frac{\sqrt{F} + \sqrt{\overline{F}}}{2}\right)^2\right) = \mathbf{Var}(F) + \frac{1}{2}F'(1)^2,$$

et donc que

$$\mathbf{E}[X_n] = O(1)$$

et

$$\mathbf{Var}[X_n] = -\left(\mathbf{Var}(F) + \frac{1}{2}F'(1)^2\right)n + O(1).$$

□

Nombres de Narayana

L'expression exacte de la fonction génératrice de probabilité des X_n fait intervenir les nombres de Narayana [98, 134]

$$N_{nk} = \frac{1}{n}\binom{n}{k}\binom{n}{k-1},$$

qui comptent le nombre d'arbres binaires avec n nœuds et k feuilles gauches (et donc $n+1-k$ feuilles droites), en bijection avec les mots de Dyck de taille n et k pics. En effet, par la formule d'inversion de Lagrange,

$$\begin{aligned}[z^n]H &= [z^n]z(F+H)(\overline{F}+H) = \frac{[x^{n-1}]}{n}(F+x)^n(\overline{F}+x)^n \\ &= \frac{1}{n}\sum_{k=0}^{n}\binom{n}{k}F^k\binom{n}{k-1}\overline{F}^{n-k+1} = \overline{F}^{n+1}\sum_{k=0}^{n}N_{nk}\left(\frac{F}{\overline{F}}\right)^k.\end{aligned}$$

Ceci montre qu'évaluer des arbres binaires avec opérateurs "$-$" et la distribution $F(u)$ aux feuilles donne la même statistique que l'évaluation des arbres binaires avec opérateurs "$+$" et distribution $F(u)$ ($\overline{F}(u)$, respectivement) aux feuilles gauches (droites, respectivement). En effet, cela peut se voir directement : la fonction génératrice bivariée pour cette dernière classe d'arbres vérifie visiblement la même équation que H (voir le début de la preuve de Prop. 1.2).

La restriction à $I = \{1\}$, pour laquelle $F^{n+1}(1) = 1$, va nous permettre de montrer plus clairement d'où proviennent les nombres de Narayana. On a la proposition :

Proposition 1.3 *Si v est défini par :*

$$v(t) = \begin{cases} 1 & \text{si } t = \boxed{1} \\ v(t_1) - v(t_2) & \text{si } t = \overset{-}{\underset{t_1 \quad t_2}{/\ \backslash}} \end{cases},$$

alors pour tout $n \geqslant 1, j \in \{1, \cdots, n\}$, le nombre $w_{n,2j-(n+1)}$ d'arbres de taille n et de valuation $2j-(n+1)$ est égal à N_{nj}.

Une bijection combinatoire permet de mieux expliquer l'apparition des nombres de Narayana. Premièrement, il est connu que les arbres généraux avec $n+1$ nœuds et j feuilles sont en bijection avec les arbres binaires à n nœuds et j feuilles gauches, suivant la correspondance de Knuth. Deuxièmement, un arbre binaire à n nœuds et j feuilles gauches possède $n+1-j$ feuilles droites. Si on remarque que $j-(n+1-j) = 2j-(n+1)$, et que la valuation k de tout arbre est de la forme $k = 2j-(n+1)$ avec $j \in \{1, \cdots, n\}$, la proposition ci-dessus peut se réécrire d'une manière combinatoire :

Proposition 1.4 *Il existe une bijection entre les arbres binaires de taille n avec opérateurs "$-$" et opérandes 1, de valuation $2j-(n+1)$, et les arbres binaires de taille n dont la différence entre le nombre de feuilles gauches et le nombre de feuilles droites vaut j.*

Démonstration. [Preuve combinatoire.] Introduisons la notion de *symétrique* d'un arbre. Si t est réduit à un simple nœud, son symétrique est t lui-même. Si $t = \circ(t_1, t_2)$, son symétrique $s(t)$ est $\circ(s(t_2), s(t_1))$.

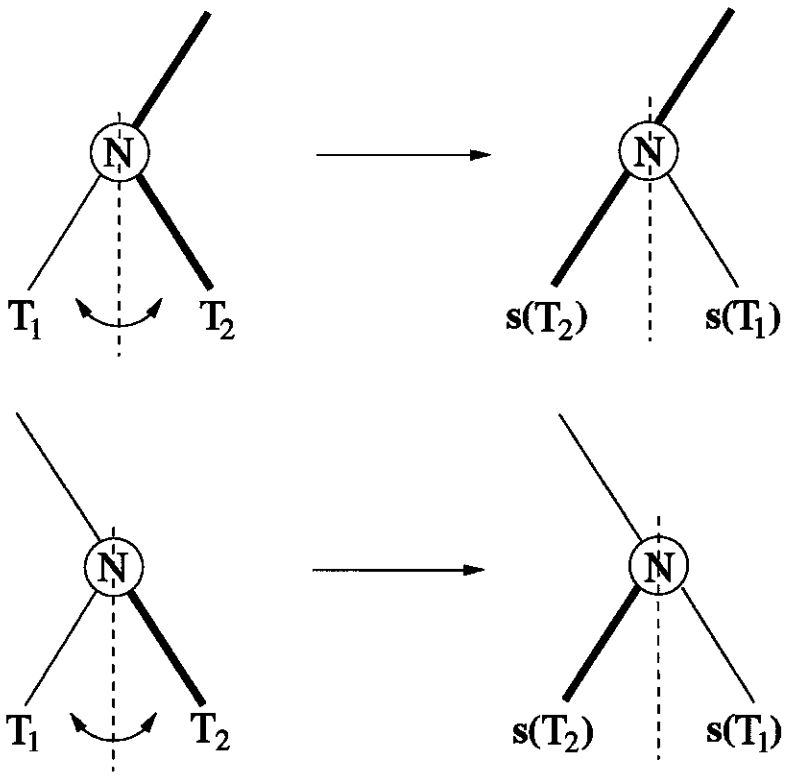
Soit t un arbre de valuation k. Appelons *arête positive* une arête telle que la valuation du sous-arbre correspondant sera comptée positivement à la racine de l'arbre, et *arête négative* une arête non positive (donc telle que la valuation du sous-arbre correspondant sera comptée négativement à la racine de l'arbre). Nous transformons t en un arbre dont la différence entre le nombre de feuilles gauches et le nombre de feuilles droites vaut k, en utilisant l'algorithme bijectif suivant, qui transforme récursivement une arête positive en une arête négative par un parcours préfixe de l'arbre (modifié au cours du parcours par l'algorithme) :

- N ← Racine.
- Répéter
  1. Visiter N.
  2. Si l'arête droite partant de N est positive, alors hanger le sous-arbre enraciné en N en son symétrique.
  3. N ← prochain nœud dans la visite préfixe de l'arbre.

  Jusqu'à (N = dernier nœud).

Bien que l'arbre soit modifié par l'algorithme, le parcours préfixe reste bien défini car les seuls les nœuds non encore rencontrés lors du parcours peuvent être déplacés par l'algorithme. □

Les figures ci-dessous illustrent les cas où l'on doit changer un sous-arbre en son symétrique. dans l'étape 2 de l'algorithme ci-dessus. Les arêtes positives sont mises en gras.



1.3.2 Arbres à opérateurs "+","−" et opérandes dans un ensemble fini I

Nous considérons maintenant des expressions plus réalistes avec à la fois des "+" et des "−", pondérés par des probabilités. Soit $\mathcal{T}((+,\alpha),(-,\beta),I)$ l'ensemble pondéré des arbres binaires de Catalan où, d'une part, les opérateurs sont égaux à "+" avec probabilité α et à "−" avec probabilité $\beta = 1-\alpha$, et d'autre part les opérandes sont choisis au hasard dans un ensemble fini d'entiers pondérés I. On notera encore $F(z) = \sum_k f_k z^k$ la série génératrice des probabilités des valeurs des opérandes.

Avant de déterminer la distribution limite, nous allons d'abord calculer la moyenne et l'espérance de la valuation des arbres considérés, en utilisant de l'analyse de singularité univariée sur les séries génératrices des moments factoriels d'ordre 1 et 2. En effet, les calculs directs sur les séries bivariées sont plus difficiles.

Proposition 1.5 *Soit X_n la variable aléatoire égale à la valuation d'un arbre de taille n tiré au hasard dans $\mathcal{T}((+,\alpha),(-,\beta),I)$ avec $\alpha < 1$. Alors*

$$\mathbf{E}[X_n] = \frac{\alpha F'(1)}{2(1-\alpha)^2} + O\left(\frac{1}{n}\right), \tag{1.8}$$

$$\mathbf{Var}[X_n] = n\left(\mathbf{Var}(F) + \frac{1+2\alpha(1-\alpha)}{2(1-\alpha)^2}F'(1)^2\right) + O(1). \tag{1.9}$$

Démonstration. Nous appellerons poids d'un arbre sa probabilité d'apparition conditionnée par la taille de cet arbre. Ainsi, si t possède $U(t)$ opérateurs "+", $|t|$ nœuds internes et $N_i(t)$ opérandes i, alors son poids est donné par l'expression

$$\text{poids}(t) = \alpha^{U(t)}\beta^{|t|-U(t)}\prod_{i\in I} f_i^{N_i(t)}.$$

Soit W la série génératrice bivariée définie par

$$W(z,u) = \sum_{t\in\mathcal{T}} \text{poids}(t) z^{|t|} u^{v(t)},$$

et posons $\overline{W}(z,u) = W(z,u^{-1})$. Comme l'opérateur "+" apparaît avec probabilité α et l'opérateur "−" avec probabilité β, $W(z,u)$ vérifie

$$\begin{aligned}
W(z,u) &= \sum_{t\in\mathcal{T}} \text{poids}(t) z^{|t|} u^{v(t)} \\
&= F + \sum_{(t_1,t_2)\in\mathcal{T}^2} \text{poids}(\oplus(t_1,t_2)) z^{1+|t_1|+|t_2|} u^{v(t_1)+v(t_2)} \\
&\quad + \sum_{(t_1,t_2)\in\mathcal{T}^2} \text{poids}(\ominus(t_1,t_2)) z^{1+|t_1|+|t_2|} u^{v(t_1)-v(t_2)} \\
&= F + \alpha z \sum_{t_1\in\mathcal{T}} \text{poids}(t_1) z^{|t_1|} u^{v(t_1)} \sum_{t_2\in\mathcal{T}} \text{poids}(t_2) z^{|t_2|} u^{v(t_2)} \\
&\quad + \beta z \sum_{t_1\in\mathcal{T}} \text{poids}(t_1) z^{|t_1|} u^{v(t_1)} \sum_{t_2\in\mathcal{T}} \text{poids}(t_2) z^{|t_2|} u^{-v(t_2)} \\
W(z,u) &= F + \alpha z W^2 + \beta z W \overline{W}.
\end{aligned} \tag{1.10}$$

En dérivant la relation ci-dessus par rapport à u et en faisant $u = 1$ on trouve (en notant que $\left(W\left(z, \frac{1}{u}\right)\right)_u = -\frac{1}{u^2} W_u\left(z, \frac{1}{u}\right)$ et donc qu'en $u = 1$ $\left(W\left(z, \frac{1}{u}\right)\right)_u\Big|_{u=1} = -W_u(z, 1)$),

$$\begin{aligned} W_u &= F_u + 2\alpha z W W_u \\ &= \frac{F_u(1 - \alpha + \alpha\sqrt{1-4z})}{1 - 2\alpha + 4\alpha^2 z}, \end{aligned}$$

d'où

$$[z^n]W_u(z, 1) = \frac{\alpha F_u}{2(1-\alpha)^2} \frac{4^n}{\sqrt{\pi n^3}} \left(1 + O\left(\frac{1}{n}\right)\right)$$

par analyse de singularité, et en utilisant

$$\mathbf{E}[X_n] = \frac{[z^n]W_u(z, 1)}{[z^n]W(z, 1)}.$$

on obtient le relation annoncée (1.8)

$$\mathbf{E}[X_n] = \frac{\alpha F'(1)}{2(1-\alpha)^2} + O\left(\frac{1}{n}\right).$$

En dérivant (1.10) une deuxième fois par rapport à u et en faisant $u = 1$ on trouve

$$\begin{aligned} (1 - 2zW)W_{u^2} &= F_{u^2} + 2(\alpha - \beta) z W_u^2 + 2\beta z W W_u \\ \sqrt{1-4z} W_{u^2} &= F_{u^2} + 2(\alpha - \beta) z \left(\frac{F_u(1 - \alpha + \alpha\sqrt{1-4z})}{1 - 2\alpha + 4\alpha^2 z}\right)^2 \\ &\quad + 2\beta z \frac{1 - \sqrt{1-4z}}{2z} \frac{F_u(1 - \alpha + \alpha\sqrt{1-4z})}{1 - 2\alpha + 4\alpha^2 z}, \end{aligned}$$

donc

$$\begin{aligned} [z^n]W_{u^2} &= \left([z^n] \frac{F_{u^2} + \frac{2(\alpha-\beta) z F_u^2 (1-\alpha)^2}{(1-\alpha)^4} + \frac{\beta F_u (1-\alpha)}{(1-\alpha)^2}}{\sqrt{1-4z}}\right) \left(1 + O\left(\frac{1}{n}\right)\right) \\ &= \left(F_{u^2} + F_u - F_u^2 + \frac{1 + 2\alpha(1-\alpha)}{2(1-\alpha)^2} F_u^2\right) \frac{4^n}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right)\right), \end{aligned}$$

et le moment factoriel d'ordre 2 est donné asymptotiquement par

$$\frac{[z^n]W_{u^2}(z, 1)}{[z^n]W(z, 1)} = \left(F_{u^2} + F_u - F_u^2 + \frac{1 + 2\alpha(1-\alpha)}{2(1-\alpha)^2} F_u^2\right)(n + O(1)).$$

Or $\mathbf{E}[X_n]$ et $(\mathbf{E}[X_n])^2$ sont en $O(1)$, donc on a la même expression asymptotique pour $\mathbf{Var}[X_n]$. □

Proposition 1.6 *Soit X_n la variable aléatoire égale à la valuation d'un arbre de taille n tiré au hasard dans $\mathcal{T}((+, \alpha), (-, \beta), I)$. Alors*

$$\frac{X_n}{\sqrt{n\left(\mathbf{Var}(F) + \frac{1+2\alpha(1-\alpha)}{2(1-\alpha)^2} F'(1)^2\right)}}$$

converge en distribution vers $\mathcal{N}(0, 1)$.

Démonstration. En remplaçant u par $1/u$ dans l'équation (1.10), on obtient une autre équation sur $\overline{W}$ et W, et les deux équations fonctionnelles satisfaites par W et $\overline{W}$ peuvent se réécrire sous la forme du système d'équations fonctionnelles

$$\begin{cases} W &= F + \alpha z W^2 + \beta z W \overline{W} \\ \overline{W} &= \overline{F} + \alpha z \overline{W}^2 + \beta z W \overline{W} \end{cases},$$

soit, si nous définissons $y_1 = W$ et $y_2 = \overline{W}$:

$$\begin{cases} y_1 &= F + \alpha z y_1^2 + \beta z y_1 y_2 \\ y_2 &= \overline{F} + \alpha z y_2^2 + \beta z y_1 y_2 \end{cases},$$

sous la forme

$$\mathbf{y} = G(z, u, \mathbf{y})$$

où $\mathbf{y} = (y_1, y_2)$, et y_1 et y_2 sont des séries holomorphes à coefficients positifs, et G est un système d'équations algébriques dont le graphe de dépendances est connexe.

Soit $\rho(u)$ la singularité de y_1. Selon [46], il suffit de montrer que

$$\mathbf{Var}\left(\frac{\rho(1)}{\rho(u)}\right) > 0, \text{ i.e. } \rho''(1) + \rho'(1) - \rho'(1)^2 < 0,$$

pour établir que la loi limite de X_n est gaussienne, où on sait que

$$\mathbf{Var}\left(\frac{\rho(1)}{\rho(u)}\right) = \underline{\lim}_{n\to\infty} \frac{\mathbf{Var}(X_n)}{n}.$$

Or les calculs précédents nous assurent cette condition avec

$$\mathbf{Var}\left(\frac{\rho(1)}{\rho(u)}\right) = \mathbf{Var}(F) + \frac{1 + 2\alpha(1-\alpha)}{2(1-\alpha)^2} F'(1)^2 > 0,$$

d'où le résultat annoncé. □

Il est théoriquement possible [46] de retrouver $\rho'(1)$ et $\rho''(1)$ à partir de

$$\begin{cases} \mathbf{y} &= G(z, u, \mathbf{y}) \\ D(z, u, \mathbf{y}) &= 0 \end{cases}, \tag{1.10}$$

où

$$\begin{aligned} D(z, u, \mathbf{y}) &= \text{dét}(Id - G_y(z, u, \mathbf{y})) \\ &= (1 - z(2\alpha y_1 + \beta y_2))(1 - z(2\alpha y_2 + \beta y_1)) - \beta^2 z^2 y_1 y_2. \end{aligned}$$

Pour ce faire on dérive le système une et deux fois par rapport à u, ce qui nous donne deux nouveaux systèmes d'équations. En faisant $u = 1$ dans ces deux derniers systèmes et dans (1.10), on obtient un système de six équations qu'il ne reste plus qu'à résoudre. Les calculs qui en découlent deviennent cependant plus compliqués, d'où l'usage précédent d'analyse de singularité univariée pour déterminer les deux premiers moments de notre variable aléatoire.

Notons que le résultat reste valide même avec un ensemble infini d'entiers, tant que la fonction génératrice de probabilité $u \mapsto F(u)$ reste analytique au voisinage de 1.

Il est aussi intéressant de remarquer que le travail ci-dessus est valide pour $\alpha = 0$ mais pas pour $\alpha = 1$, ce qui correspond au cas avec seulement des plus. Concluons en récapitulant comment l'espérance et la variance de X_n varient avec α. Elles dépendent de l'espérance $F'(1)$ et plus généralement de la distribution des entiers aux nœuds externes de la manière suivante :

- Si $F'(1) = 0$, alors bien sûr $E[X_n] = 0$, et nous avons obtenu que la variance de X_n est asymptotiquement égale à $n\,\mathbf{Var}(F)$ pour tout α.
- Si $F'(1)$ est non nul, alors nous avons $E[X_n] = \Theta(1)$ pour $\alpha < 1$, $E[X_n] = (n+1)F'(1)$ pour $\alpha = 1$, et par conséquent il n'est pas surprenant que la variance de la distribution limite, vue comme fonction de α, ne soit pas continue en $\alpha = 1$.

1.4 Arbres $+, \times$

Il est possible de calculer les premiers moments de la valuation des arbres avec opérateurs $+$ et $\times$ qui apparaissent dans [12]. Nous notons à nouveau I l'ensemble des opérandes possibles, à valeurs dans $\mathbb{N}$, et $F(z) = \sum_{i \in I} p_i z^i$ la fonction génératrice de probabilités des opérandes placés aux nœuds externes. Nous notons α et β les probabilités d'apparition respectives des opérateurs addition et multiplication. À chaque arbre t est associé son poids poids(t), qui représente sa probabilité d'apparition conditionnée par sa taille.

1.4.1 Séries génératrices des puissances des moments

Proposition 1.7 *Soit $\mathcal{T}$ la classe des arbres pondérés à opérateurs $+$ et $\times$ et opérandes dans I. Soit $f^{(m)}$ le moment d'ordre m de la distribution des opérandes, autrement dit*

$$f^{(m)} = \sum_{k \in I} p_k k^m.$$

Alors la série génératrice des puissances d'ordre m des valuations des arbres d'une taille donnée vérifie :

$$M_m(z) = f^{(m)} + \alpha z \sum_{p=0}^{m} \binom{m}{p} M_p M_{m-p} + \beta z M_m^2.$$

Démonstration. La classe des arbres étiquetés pondérés est donnée par :

$$\begin{aligned} \mathcal{T} &= \textstyle\sum_{t \in \mathcal{T}} \text{poids}(t)\, t \\ &= \textstyle\sum_{i \in I} p_i \boxed{\text{i}} \;+\; \alpha \;\; \overset{+}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{\diagup \quad \diagdown}} \;+\; \beta \;\; \overset{\times}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{\diagup \quad \diagdown}}, \end{aligned}$$

donc la série génératrice correspondante T vérifie :

$$T(z) = 1 + zT(z)^2 = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

La classe des arbres étiquetés comptés par valuations est donnée par :

$$\begin{aligned} \mathcal{M}_1 &= \sum_{T \in \mathcal{T}} v(T) T \\ &= \textstyle\sum_{i \in I} p_i\, i \boxed{\text{i}} \;+\; \alpha \;\; \overset{+}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{\diagup \quad \diagdown}} (v(\mathcal{T}_1) + v(\mathcal{T}_2)) \;+\; \beta \;\; \overset{\times}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{\diagup \quad \diagdown}} (v(\mathcal{T}_1).v(\mathcal{T}_2)) \end{aligned}$$

d'où, pour la série génératrice correspondante, en remarquant $M_0 = T$:

$$M_1 = f^{(1)} + 2\alpha z T M_1 + \beta z M_1^2.$$

Pour les moments d'ordre 2 on trouve :

$$\mathcal{M}_2 = \sum_{t\in\mathcal{T}} v(t)^2 t$$

$$= \boxed{1} + \alpha \; \overset{+}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{/\ \backslash}} (v(\mathcal{T}_1)+v(\mathcal{T}_2))^2 + \beta \; \overset{\times}{\underset{\mathcal{T}_1 \quad \mathcal{T}_2}{/\ \backslash}} (v(\mathcal{T}_1).v(\mathcal{T}_2))^2 \ ,$$

d'où, pour la série génératrice correspondante :

$$M_2 = f^{(2)} + 2\alpha z T M_2 + 2z M_1^2 + \beta z M_2^2.$$

De manière plus générale, on établit avec des calculs similaires que la fonction génératrice $M_m(z) = \sum_{n\geqslant 0} \sum_{T,|T|=n} v(T)^m z^n$ des moments d'ordre m vérifie :

$$M_m(z) = f^{(m)} + \alpha z \sum_{p=0}^{m} \binom{m}{p} M_p M_{m-p} + \beta z M_m^2.$$

□

Ce résultat nous permet de calculer une série M_m à l'aide des séries précédentes M_k pour $k < m$, puisqu'en fait la récurrence n'est rien d'autre qu'une équation du second degré en M_m, ce que nous résumons en le corollaire

Corollaire 1.1 *Avec les mêmes notations, pour $m \geqslant 1$ la série génératrice des puissances d'ordre m des valuations des arbres d'une taille donnée vérifie, pour $0 < \alpha < 1$:*

$$M_m(z) = \frac{1 - 2\alpha z T - \sqrt{(1-2\alpha z T)^2 - 4z\left(f^{(m)} + \alpha z \sum_{p=1}^{m-1} \binom{m}{p} M_p M_{m-p}\right)}}{2\beta z}. \tag{1.11}$$

1.4.2 Asymptotique des moments

Le moment d'ordre m des valuations des arbres de taille n est donné par

$$\frac{[z^n]M_m(z)}{[z^n]T(z)},$$

dont on peut calculer l'asymptotique en effectuant une analyse de singularité des séries génératrices T et M_m.

Si I se compose d'éléments strictement positifs, il est possible de déterminer la nature de l'asymptotique des coefficients de M_m.

Proposition 1.8 *Si les éléments de I sont strictement positifs, alors pour tout m il existe $\lambda_m > 0$ et ρ_m tels que*

$$[z^n]M_m(z) \underset{n\to\infty}{\sim} \lambda_m \frac{\rho_m^n}{\sqrt{\pi n^3}}, \tag{1.12}$$

et la suite $(\rho_m)_{m\in\mathbb{N}}$ est strictement décroissante.

Démonstration. Soit $\mathcal{P}(m)$ la propriété composée de (1.12) et de la stricte décroissance de la suite $(\rho_m)_{m\in[\![0,m-1]\!]}$.

Montrons cette propriété par récurrence. Pour $m = 0$, on a

$$M_0(z) = T(z) = \frac{1-\sqrt{1-4z}}{2z},$$

et on retrouve l'asymptotique connue des nombres de Catalan, ce qui montre $\mathcal{P}(0)$.

Pour $m = 1$, on a

$$M_1(z) = \frac{1-2\alpha zT-\sqrt{(1-2\alpha zT)^2-4\beta zF'(1)}}{2\beta z}.$$

L'expression

$$R_1(z) = (1-2\alpha zT)^2 - 4\beta zF'(1)$$

est visiblement analytique pour $z \leqslant \rho_0 = \frac{1}{4}$. En $z = \frac{1}{4}$, sachant que $T(\frac{1}{4}) = 2$ on a

$$R_1\left(\frac{1}{4}\right) = \left(1-2\alpha\cdot\frac{1}{4}\cdot 2\right)^2 - 4\beta\cdot\frac{1}{4}f^{(1)} = \beta^2 - \beta f^{(1)} < 0.$$

Comme d'autre part $R(0) = 1$ est strictement positif, on en déduit que R admet un zéro dans $]0, \frac{1}{4}[$. Soit ρ_1 le plus petit zéro de R. C'est une singularité de type quadratique pour M_1 qui vérifie $\rho_1 < \rho_0$, d'où la propriété $\mathcal{P}(1)$.

Soit maintenant m plus grand que 2, et supposons $\mathcal{P}(k)$ montré pour $k \geqslant m-1$. De l'expression générale (1.11) des M_m, on déduit que la singularité dominante de ρ_m est soit égale à la plus petite des singularités dominantes des M_k pour $k < m$, c'est-à-dire ρ_{m-1}, soit au plus petit réel z positif tel que l'expression

$$R_m(z) = (1-2\alpha zT)^2 - 4z\left(f^{(m)} + \alpha z\sum_{p=1}^{m-1}\binom{m}{p}M_p(z)M_{m-p}(z)\right)$$

soit nulle.

L'expression R_m est visiblement analytique au moins jusqu'à ρ_{m-1}. Or on sait que ρ_{m-1} est racine de R_{m-1}, ce qui s'écrit

$$4\rho_{m-1}\left(f^{(m-1)} + \alpha\rho_{m-1}\sum_{p=1}^{m-2}\binom{m-1}{p}M_p(\rho_{m-1})M_{m-1-p}(\rho_{m-1})\right) = (\beta+\alpha\sqrt{1-4\rho_{m-1}})^2.$$

Comme on travaille avec des opérandes plus grands que 1, on a

$$f^{(m)} = \sum_{k\in I} p_k k^m \geqslant f^{(m-1)} = \sum_{k\in I} p_k k^m.$$

Le fait d'avoir des opérandes plus grands que 1 et des opérateurs $+$ et $\times$ implique également qu'on a des valuations plus grandes que 1, et, tant que z réel positif reste inférieur aux rayons de convergence des séries considérées, si $p < q$,

$$M_p(z) = \sum_T v(T)^p z^{|T|} \leqslant M_q(z) = \sum_T v(T)^q z^{|T|}.$$

Pour tous $m \geqslant 1$ et p entiers, il est clair aussi qu'on a l'égalité suivante :

$$\binom{m-1}{p} \geqslant \binom{m}{p}.$$

De ces diverses inégalités on déduit que la quantité

$$4z\left(f^{(m)} + \alpha z \sum_{p=1}^{m-1} \binom{m}{p} M_p(z) M_{m-p}(z)\right)$$

est strictement plus grande que

$$4z\left(f^{(m-1)} + \alpha z \sum_{p=1}^{m-2} \binom{m-1}{p} M_p(z) M_{m-1-p}(z)\right)$$

pour $z \in [0, \rho_{m-1}]$. Par conséquent, $R_m(\rho_{m-1})$ est strictement négatif, et d'autre part $R_m(0)$ vaut 1 et est strictement positif, donc R_m admet un zéro strictement compris entre 0 et ρ_{m-1}. Soit ρ_m le plus petit de ces zéros. C'est une singularité de type quadratique pour M_m qui vérifie $\rho_m < \rho_{m-1}$, d'où la propriété $\mathcal{P}(m)$. □

Les résultats ci-dessus ne subsistent pas forcément si on travaille également avec des opérandes strictement plus petits que 1. Un travail futur consisterait à établir une typologie des comportements de M_m suivant ces opérandes.

Calcul numériques pour $I = \{1\}$

En effectuant une analyse de singularité sur les premiers M_m, on trouve numériquement les résultats suivants :

- $[z^n]M_0(z) \sim \frac{4^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_1(z) \sim 3\sqrt{3}\frac{4,5^n}{\sqrt{\pi n^{3/2}}} \sim 5,196152420\frac{4,5^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_2(z) \sim 7,163798465\frac{(5,373242625)^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_3(z) \sim 9,711750970\frac{(6,669033195)^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_4(z) \sim 13,15596181\frac{(8,538560555)^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_5(z) \sim 17,91555400\frac{(11,21106734)^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_6(z) \sim 24,57004515\frac{(15,01857573)^n}{\sqrt{\pi n^{3/2}}}$
- $[z^n]M_7(z) \sim 33,934488649\frac{(20,43733673)^n}{\sqrt{\pi n^{3/2}}}$

1.4.3 Conséquences sur la distribution limite

Dans le cas où nous avons une explosion exponentielle des moments, on ne peut pas espérer avoir une distribution limite sympathique à partir d'une normalisation du type

$$\frac{X_n - \mathbf{E}[X_n]}{\mathbf{Var}[X_n]},$$

puisque la fonction caractéristique limite aurait un rayon de convergence nul.

1.5 Conclusion

Suivant les opérateurs des expressions arithmétiques tirés au hasard, la distribution limite du résultat de l'expression varie beaucoup. Avec des "min" et des "+", et des opérandes positifs, on obtient une limite discrète. La méthodologie s'applique aussi au paramètre 2-bud qui fait intervenir l'opérateur min. Il semblerait que des paramètres extrémaux avec des opérations de type minimisation mène à des limites discrètes, ce qui n'est pas très étonnant lorsqu'on sait qu'un arbre de Catalan a une chance sur deux d'avoir un sous-arbre réduit à une feuille.

Avec des opérateurs "+" et des "−", on obtient une limite gaussienne. Passer à travers une structure algébrique qu'est l'arbre semble ne pas changer le phénomène de théorème central limite. Cependant, lorsque la variance de la distribution F des opérandes aux nœuds externes n'est pas nulle, il se passe un phénomène intéressant : si α est la probabilité d'avoir un $+$ aux nœuds internes, alors on a une espérance asymptotiquement bornée, et une variance de l'ordre de n fois $\mathbf{Var}(F) + \frac{1+2\alpha(1-\alpha)}{2(1-\alpha)^2} F'(1)^2$ au lieu de seulement n fois $\mathbf{Var}(F)$ si on ne faisait que des additions.

Avec des "+" et des "×", on observe une explosion exponentielle des moments.

Chapitre 2

Valuations arithmétiques sur des arbres tirés suivant la distribution des arbres binaires de recherche

Dans ce chapitre, nous effectuons une étude similaire à l'étude précédente sur des arbres de Catalan pour des arbres avec opérateurs "+" et "−", en tirant cette fois les arbres selon la distribution des arbres binaires de recherche. Pour la définition des arbres binaires de recherche, nous renvoyons à [82], où l'on voit que ces arbres ont la même distribution que les arbres de permutation.

Les arbres binaires de recherche sont surtout utilisés pour étudier le tri rapide : un arbre binaire de recherche de taille n donnée représente l'historique du tri rapide appliqué à une permutation de n entiers distincts. Le coût du tri rapide correspond alors à la longueur de cheminement interne de l'arbre, c'est-à-dire à la somme des distances des nœuds à la racine. La distribution limite du coût du tri rapide ou quicksort est toujours inconnue et fait encore l'objet d'une étude intensive [32, 44, 66, 67, 68, 69, 105, 127, 159, 172, 181, 201]. Le quicksort lui-même fait l'objet de diverses recherches en vue d'être optimisé [57, 158, 155].

Les arbres binaires de recherche obéissent à une statistique différente des arbres de Catalan et en particulier sont globalement plus équilibrés. Alors que tous les arbres de Catalan d'une taille n donnée sont équiprobables, la probabilité de tirer un arbre T est donnée par [132, 154] :

$$P(T) = \begin{cases} 1 & \text{si } T = \square \\ \dfrac{P(T^l) \cdot P(T^r)}{1 + |T^l| + |T^r|} & \text{si } T = \circ(T^l, T^r). \end{cases} \tag{2.1}$$

Un même paramètre sur des arbres de Catalan ou de recherche suivra donc des statistiques différentes, et aura donc une distribution limite différente, lorsque n tend vers l'infini. Ainsi, la distribution limite de la longueur de cheminement des arbres de Catalan est connue, et diffère de la distribution limite de la longueur de cheminement des arbres binaires de recherche.

Divers paramètres sont étudiés sur les arbres binaires de recherche [5, 28, 58, 154, 156, 157], et souvent par intérêt essentiellement mathématique on peut constater des différences entre les distributions limite, sinon en caractérisant leur nature, au moins en comparant leur moyenne. Parmi les paramètres étudiés, on peut citer en particulier la recherche de motifs [73], le nombre de Strahler [43], la hauteur [41].

Nous verrons que notre valuation arithmétique avec opérateurs + et −, et toujours opérandes choisis parmi un ensemble fini d'entiers, possède une espérance asymptotiquement différente de

celle du même paramètre sur les arbres de Catalan. Pour notre paramètre d'intérêt, on notera encore α la probabilité de mettre un + aux nœuds internes et $\beta = 1 - \alpha$ la probabilité de mettre un −. Ainsi, si l'arbre d'expression t possède $U(t)$ opérateurs "+", $|t|$ noeuds internes et $N_i(t)$ opérandes i, alors son poids est donné par l'expression

$$\text{poids}(t) = \alpha^{U(t)}\beta^{|t|-U(t)} \prod_{i \in I} f_i^{N_i(t)} \cdot P(T),$$

avec $P(T)$ défini comme dans (2.1).

Nous verrons que la nature de la distribution limite varie en fonction de α et de la distribution des opérandes. La loi limite sera le plus souvent gaussienne, mais dès que la moyenne des opérandes est non nulle et que α est strictement compris entre $\frac{3}{4}$ et 1, nous avons affaire à une distribution limite exotique asymétrique, que nous caractérisons à l'aide d'une équation fonctionnelle sur la fonction caractéristique par exemple.

Comme nous le verrons dans le chapitre suivant consacré aux martingales, ce paramètre est d'autant plus intéressant qu'il suit une récurrence rappelant celle qu'on peut trouver avec le polynôme de niveau (chapitre 3 remarque 3.1).

2.1 Arbres +, − de recherche

On met un + avec probabilité α, et un − avec probabilité $\beta = 1 - \alpha$. On suppose que la distribution de chaque opérande suit la même loi, de manière indépendante des autres opérandes, et qu'elle est donnée par une série génératrice F analytique en 1. On étudie la valuation X_n des ABR ainsi étiquetés.

2.1.1 Moyenne et variance

Nous calculons la moyenne et la variance de la distribution limite en utilisant des fonctions génératrices.

Proposition 2.1 *Posons $f' = F'(1)$ et $f'' = F''(1)$. Soient*

$$C_\alpha = \frac{\alpha - 1/2}{\alpha - 3/4}\ ; \qquad \lambda_0 = f'' - C_\alpha f'^2 + f'.$$

Supposons F non dégénérée, i.e. de variance non nulle. Alors l'espérance μ_n de X_n vaut

$$\mu_n = f' \frac{\Gamma(n + 2\alpha)}{\Gamma(2\alpha) n!},$$

avec, si $f' \neq 0$,

$$\mu_n \sim \frac{n^{2\alpha-1}}{\Gamma(2\alpha)} f',$$

et la variance σ_n^2 dépend des valeurs de f' et de α de la manière qui suit :

Si $0 \leqslant \alpha < 3/4$ ou $\alpha = 1$, ou si $f' = 0$, alors :

$$\sigma_n^2 \sim \lambda_0 n.$$

Si $\alpha = 3/4$ et $f' \neq 0$, alors :

$$\sigma_n^2 \sim f'^2 n \log n.$$

Si $3/4 < \alpha < 1$ *et* $f' \neq 0$, *alors :*

$$\sigma_n^2 \sim \left(\frac{C_\alpha}{\Gamma(4\alpha - 1)} - \frac{1}{\Gamma(2\alpha)^2} \right) f'^2 n^{4\alpha - 2}.$$

Démonstration. Notons à nouveau $v(t)$ la valuation de t.

$$W(z, u) = \sum_t \text{poids}(t) z^{|t|} u^{v(t)}.$$

En utilisant l'expression du poids de chaque arbre t en fonction de celles de ses sous-arbres, s'ils ne sont pas vides, ou bien en procédant plus formellement comme dans [82], on obtient facilement le système d'équations

$$\begin{cases} W_z &= \alpha W^2 + \beta W \overline{W}, \\ \overline{W}_z &= \alpha \overline{W}^2 + \beta W \overline{W}, \end{cases} \tag{2.2}$$

avec les conditions limites

$$\begin{cases} W(z, 1) = \dfrac{1}{1 - z} \quad , & W_u(0, u) = F', \\ W(0, u) = F \quad , & W_u^2(0, u) = F''. \end{cases}$$

Pour calculer moyenne et variance, nous allons dériver (2.2) une et deux fois par rapport à u et faire $u = 1$.

En dérivant une première fois, on trouve

$$W_{zu} = 2\alpha W W_u + \beta W_u \overline{W} + \beta W \overline{W}_u,$$

qui devient en $u = 1$ (car $\overline{W}_u = -\frac{1}{u^2} W_u$) :

$$\begin{aligned} W_{zu} &= 2\alpha W W_u \\ &= \frac{2\alpha}{1 - z} W_u. \end{aligned}$$

Résolvons cette équation. On a

$$\frac{W_{zu}(z, 1)}{W_u(z, 1)} = \frac{2\alpha}{1 - z},$$

et il existe c tel que

$$W_u(z, 1) = e^c e^{2\alpha \log(1 - z)}.$$

Or $W_u(0, 1) = f'$ donc $e^c = f'$ et : $W_u(z, 1) = \dfrac{f'}{(1 - z)^{2\alpha}}$. On en déduit :

$$[z^n] W_u(z, 1) = f' \frac{\Gamma(n + 2\alpha)}{\Gamma(2\alpha) n!}.$$

Pour $\alpha \neq 0$, on a l'asymptotique

$$[z^n] W_u(z, 1) \sim \frac{n^{2\alpha - 1}}{\Gamma(2\alpha)}.$$

En dérivant (2.2) une seconde fois, on obtient

$$W_{zu^2} = 2\alpha\left(W_u^2 + WW_{u^2}\right)\left(W_{u^2}\overline{W} + W_u\overline{W}_u + \overline{W}_{u^2}\right),$$

soit, en $u = 1$:

$$\begin{aligned}
W_{zu^2} &= 2\alpha\frac{f'}{(1-z)^{4\alpha}} + 2\alpha\frac{W_{u^2}}{(1-z)} + \beta\frac{W_{u^2}}{(1-z)} - 2\beta\frac{f'^2}{(1-z)^{4\alpha}} + \beta\frac{W_{u^2} + 2W_u}{(1-z)} \\
&= 2\alpha\frac{f'}{(1-z)^{4\alpha}} + 2\alpha\frac{W_{u^2}}{(1-z)} + \beta\frac{W_{u^2}}{(1-z)} - 2\beta\frac{f'^2}{(1-z)^{4\alpha}} + \beta\frac{W_{u^2} + 2\frac{f'}{(1-z)^{2\alpha}}}{(1-z)} \\
&= 2(\alpha-\beta)\frac{f'}{(1-z)^{4\alpha}} + 2(\alpha+\beta)\frac{W_{u^2}}{(1-z)} + 2\beta\frac{f'}{(1-z)^{2\alpha+1}} \\
&= 2(\alpha-\beta)\frac{f'}{(1-z)^{4\alpha}} + 2\frac{W_{u^2}}{(1-z)} + 2\beta\frac{f'}{(1-z)^{2\alpha+1}}.
\end{aligned}$$

La solution de l'équation homogène s'écrit :

$$W_{u^2}(z,1) = \frac{\lambda}{(1-z)^2},$$

donc on utilise la méthode de la variation de la constante en écrivant :

$$W_{u^2}(z,1) = \frac{\lambda(z)}{(1-z)^2}.$$

Il vient :

$$W_{zu^2}(z,1) = 2\frac{\lambda(z)}{(1-z)^3} + \frac{\lambda'(z)}{(1-z)^2}.$$

Si $\alpha \neq 3/4$, on a

$$W_{u^2}(z,1) = \frac{C_\alpha f'^2}{(1-z)^{4\alpha-1}} - \frac{f'}{(1-z)^{2\alpha}} + \frac{\lambda_0}{(1-z)^2}.$$

où

$$C_\alpha = \frac{\alpha - 1/2}{\alpha - 3/4}, \qquad \lambda_0 = f'' - C_\alpha f'^2 + f'.$$

Si $\alpha = 3/4$, on a

$$W_{u^2}(z,1) = -\frac{f'^2\log(1-z)}{(1-z)^2} - \frac{f'}{(1-z)^{3/2}} + \frac{f+f''}{(1-z)^2}.$$

En effectuant une analyse de singularité sur

$$\sigma_n^2 = ([z^n]W_{u^2}(z,1) + W_u(z,1)) - ([z^n]W_u(z,1))^2,$$

on trouve les asymptotiques annoncées. □

Nous voyons donc que la variance est linéaire dans le cas $F'(1) = 0$. Si $F'(1) \neq 0$, la variance n'est linéaire que dans le cas dégénéré $\alpha = 1$, et dans le cas $\alpha < \frac{3}{4}$. Le fait de mettre une loi de moyenne non nulle aux nœuds externes introduit un déséquilibre.

2.1.2 Point fixe

Nous montrons dans ce paragraphe que pour le cas où la variance est linéaire, ou presque — nous traitons aussi le cas où la variance est en $n \log n$ —, la distribution limite est gaussienne. Nous utiliserons pour cela des théorèmes de point fixe, comme présentés dans B.

Commençons d'abord pas le cas où la variance est linéaire.

Théorème 2.1 *Soit X_n la variable aléatoire égale à la valuation de l'arbre de taille n tiré dans notre modèle. Lorsque la variance est linéaire, c'est-à-dire lorsque l'une des conditions $F'(1) = 0$, $0 \leqslant \alpha < 3/4$ ou $\alpha = 1$ est remplie, alors X_n converge en distribution vers une loi limite normale.*

Démonstration. Comme on a un opérateur soustractif, on ne peut pas utiliser la boîte à outils de l'annexe B, donc on va montrer la convergence à "la main", comme dans [180].

Il faut d'abord avoir une idée de la distribution limite, puis montrer rigoureusement la convergence vers cette distribution limite.

Soit Y une variable aléatoire telle que $\mathrm{P}[Y = 1] = \alpha = 1 - \beta = 1 - \mathrm{P}[Y = 0]$. Étant donné la définition récursive des arbres binaires de recherche, X_n vérifie l'identité

$$X_n \stackrel{d}{=} Y(A_{Z_n} + B_{n-1-Z_n}) + (1 - Y)(C_{Z_n} - D_{n-1-Z_n}) \tag{2.3}$$

avec pour tout n X_n, A_n, B_n, C_n, D_n i.i.d, et $\forall k \in \{0, \ldots, n-1\}, P[Z_n = k] = \frac{1}{n}$. Z_n représente la taille de l'arbre gauche, A_{Z_n} et C_{Z_n} la valuation de l'arbre gauche dans les cas où l'opérateur est respectivement un plus ou un moins, et B_{Z_n} et D_{Z_n} la valuation de l'arbre droit dans les cas où l'opérateur est respectivement un plus ou un moins.

La variance étant asymptotiquement linéaire, on introduit la variable normalisée $X_n^* := \frac{X_n - E[X_n]}{\sqrt{n}}$.

On a $E[X_n] \sim \dfrac{f' n^{2\alpha-1}}{\Gamma(2\alpha)}$ dans le cas $f' \neq 0$ ($E[X_n]$ est nul si $f' = 0$), et $E[X_n^2] \sim \lambda_0 n$, avec

$$C_\alpha = \frac{\alpha - 1/2}{\alpha - 3/4} \; ; \qquad \lambda_0 = f'' - C_\alpha f'^2 + f'$$

Si $f' = 0$, alors $E[X_n] = 0$ et $E[X_n^2] = f''n$. Si $\alpha < 3/4$, on a $E[X_n]/\sqrt{E[X_n^2]} \to 0$, soit $\mathrm{Var}[X_n] \sim E[X_n^2] \sim \lambda_0 n$. (Naturellement, on a bien $\lambda_0 > 0$ car $\lambda_0 = \mathrm{Var}(f) + (1 - C_\alpha) f'^2$ avec $C_\alpha < 0$ pour $\alpha < 3/4$).

Dans les deux cas ci-dessus, $X_n^* := \dfrac{X_n - E[X_n]}{\sqrt{n}} \sim \dfrac{X_n}{\sqrt{n}} \propto \dfrac{X_n}{\sqrt{\mathrm{Var}(X_n)}}$.

On définit σ_n^2 par $\sigma_n^2 = E[(X_n^*)^2]$, et σ^2 par $\sigma^2 = \lambda_0$ si $0 \leqslant \alpha < 3/4$ ou $f' = 0$, et par $\sigma^2 = f''$ si $\alpha = 1$.

Le relation (2.3) sur X_n se traduit sur X_n^* par

$$\begin{aligned} X_n^* \stackrel{d}{=} \; & Y\left(\sqrt{\frac{Z_n}{n}} A^*_{Z_n} + \sqrt{\frac{n-1-Z_n}{n}} B^*_{n-1-Z_n}\right) \\ & + (1-Y)\left(\sqrt{\frac{Z_n}{n}} C^*_{Z_n} - \sqrt{\frac{n-1-Z_n}{n}} D^*_{n-1-Z_n}\right). \end{aligned}$$

À la limite, si τ désigne la loi uniforme sur $[0, 1]$, et si on appelle X, A, B, C et D les limites respectives de (X_n^*), (A_n^*), (B_n^*), (C_n^*) et (D_n^*), on obtient

$$X \stackrel{d}{=} Y(\tau^{1/2} A + (1-\tau)^{1/2} B) + (1-Y)(\tau^{1/2} C - (1-\tau)^{1/2} D). \tag{2.4}$$

Remarquons d'abord que les lois normales, de moyenne nulle, et de variance quelconque, sont solutions de cette équation en distribution (pour le voir, il suffit de réécrire cette équation avec des fonctions caractéristiques). Si σ_n^2 est la variance de X_n^*, on peut donc raisonnablement penser que X_n^* converge vers une loi normale de variance $\sigma^2 = \lim \sigma_n^2$.

Pour montrer cette conjecture, nous allons définir des lois normales accompagnatrices N_n de même variance σ_n^2 que X_n^*, dont on sait qu'elles convergent vers $\mathcal{N}(0, \sigma^2)$. Nous montrerons alors que (N_n) et X_n^* ont même limite, ce qui établira le résultat voulu.

Nous allons utiliser un théorème de point fixe suivant la métrique

$$\zeta_r = \sup\{|Ef(X) - Ef(Y)|; f \in C^{(2)}, ||f^{(2)}||_{1/(3-r)} \leqslant 1\},$$

où r est un réel quelconque fixé strictement compris entre 2 et 3.

En effet, Z_n vérifie de bonnes propriétés :

$$\begin{aligned} E\left[\left(\frac{Z_n}{n}\right)^{r/2} + \left(\frac{n-Z_n}{n}\right)^{r/2}\right] &= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^{r/2} + \left(\frac{n-k}{n}\right)^{r/2} \\ &\underset{n\to\infty}{\longrightarrow} 2\int_0^1 x^{r/2}dx = \frac{2}{1+r/2} < 1 \end{aligned}$$

(il s'agit de sommes de Riemann), autrement dit $b_r = \overline{\lim}\ \mathbf{E}\left[\left(\frac{Z_n}{n}\right)^{r/2} + \left(\frac{n-Z_n}{n}\right)^{r/2}\right] < 1$, où b_r joue physiquement le rôle d'un coefficient de contraction, comme nous le verrons à la fin de la preuve ci-dessous.

Nous introduisons maintenant pour chaque entier n une loi normale de même variance que X_n. Nous montrons que ces lois normales vérifient des récurrences similaires à celles de X_n, qu'elles admettent une loi limite normale, et que (X_n) sont proches de ces lois normales et admet la même loi limite.

Lemme 2.1 *Soient $N_n \stackrel{d}{=} \mathcal{N}(0, \sigma_n^2)$, $a_n = \zeta_r(X_n^*, N_n)$. Soit $a = \sup a_n$. Alors : $a < \infty$.*

Démonstration. Soit

$$\kappa_r = r\int_{-\infty}^{\infty} |x|^{r-1}\,|F_X(x) - F_Y(x)|\,dx.$$

On sait qu'on a alors l'inégalité [178] :

$$\zeta_r(X,Y) \leqslant \frac{\Gamma(r-1)}{\Gamma(1+r)}\kappa_r(X,Y),$$

et par conséquent, en appliquant cette inégalité à X_n^* et N_n on trouve l'existence d'une constance $C > 0$ telle que

$$\sup a_n \leqslant C \sup_n (\mathbf{E}|X_n^*|^r + \mathbf{E}|N_n|) < \infty.$$

□

Lemme 2.2 *Posons*

$$\begin{aligned} N_n^* &:= Y\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(1)} + \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(2)}\right) \\ &\quad +(1-Y)\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(3)} - \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(4)}\right), \end{aligned}$$

où les $N_k^{(j)}$ *sont des copies indépendantes de* N_k. *On a alors* $\overline{\lim}\ \zeta_r(X_n^*, N_n^*) \leqslant ab_r < \infty$.

Démonstration.

$$\begin{aligned}
\zeta_r(X_n^*, N_n^*) &\leqslant \alpha \sum_{k=0}^{n} P[Z_n = k]\ \zeta_r\left(\left(\frac{k}{n}\right)^{1/2} N_k^{(1)} + \left(\frac{n-k}{n}\right)^{1/2} N_{n-k}^{(2)},\right. \\
&\quad \left.\left(\frac{k}{n}\right)^{1/2} X_k^* + \left(\frac{n-k}{n}\right)^{1/2} \overline{X}_{n-k}^*\right) \\
&\quad + \beta \sum_{k=0}^{n} P[Z_n = k]\ \zeta_r\left(\left(\frac{k}{n}\right)^{1/2} N_k^{(3)} + \left(\frac{n-k}{n}\right)^{1/2} N_{n-k}^{(4)},\right. \\
&\quad \left.\left(\frac{k}{n}\right)^{1/2} X_k^* + \left(\frac{n-k}{n}\right)^{1/2} \overline{X}_{n-k}^*\right) \\
&\leqslant \sum_{k=0}^{n} P[Z_n = k]\ \zeta_r(X_k^*, N_k^*) \left[\left(\frac{k}{n}\right)^{r/2} + \left(\frac{n-k}{n}\right)^{r/2}\right] \\
&\leqslant a\left[E\left(\frac{Z_n}{n}\right)^{r/2} + E\left(\frac{n-Z_n}{n}\right)^{r/2}\right].
\end{aligned}$$

Donc

$$\overline{\lim}\ \zeta_r(X_n^*, N_n^*) \leqslant ab_r < \infty.$$

□

Comme conséquence de ce lemme, on obtient, d'après la proposition B.3 de l'annexe B :

$$E[X_n^*] = E[N_n^*] = 0, \qquad E[(X_n^*)^2] = E[(N_n^*)^2] = \sigma_n^2.$$

Lemme 2.3 *Soit* $d_n = \zeta_r(N_n, N_n^*)$. *Alors* $d_n \to 0$.

Démonstration. Comme la moyenne et la variance de N_n coïncident avec celles de N_n^*, puisque l'inégalité

$$\zeta_r(N_n^*, X_n^*) < \infty$$

implique

$$E[(N_n^*)^j - (X_n^*)^j] = 0,\ j, = 1, 2,$$

toujours d'après la proposition B.3 de l'annexe B, il s'ensuit que $d_n \leqslant C\kappa_r(N_n, N_n^*)$ pour une certaine constante C. Soit $\overline{N}_0$ une variable aléatoire normale standard indépendante de Z_n et de Y. Rappelons au passage les relations

$$\mathcal{N}(0, \sigma^2) \equiv \sigma\mathcal{N}(0, 1)$$

et

$$\alpha_1\mathcal{N}(0, 1) + \alpha_2\mathcal{N}(0, 1) \equiv \sqrt{\alpha_1^2 + \alpha_2^2}\mathcal{N}(0, 1)$$

pour tous réels α_1, α_2. Alors

$$\begin{aligned}
N_n^* &= Y\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(1)} + \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(2)}\right) \\
&\quad +(1-Y)\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(3)} - \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(4)}\right) \\
&\stackrel{d}{=} Y\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(1)} + \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(2)}\right) \\
&\quad +(1-Y)\left(\sqrt{\frac{Z_n}{n}}N_{Z_n}^{(3)} + \sqrt{\frac{n-1-Z_n}{n}}N_{n-1-Z_n}^{(4)}\right) \\
&\stackrel{d}{=} Y\left(\frac{Z_n}{n}\sigma_{Z_n}^2 + \frac{n-1-Z_n}{n}\sigma_{n-1-Z_n}^2\right)^{1/2} N_0 \\
&\quad +(1-Y)\left(\frac{Z_n}{n}\sigma_{Z_n}^2 - \frac{n-1-Z_n}{n}\sigma_{n-1-Z_n}^2\right)^{1/2} \overline{N}_0 \\
&\stackrel{d}{=} \left(\frac{Z_n}{n}\sigma_{Z_n}^2 + \frac{n-1-Z_n}{n}\sigma_{n-1-Z_n}^2\right)^{1/2} N_0 \\
&=: \eta_n N_0,
\end{aligned}$$

avec

$$\eta_n \to \left(p\sigma^2 + (1-p)\sigma^2\right)^{1/2} = \sigma,$$

donc $d_n = \zeta_r(N_n, N_n^*) \to 0$. □

Avec

$$a_n = \zeta_r(X_n^*, N_n) \leqslant \zeta_r(X_n^*, N_n^*) + d_n,$$

posant $\overline{a} = \overline{\lim}\, a_n$, on obtient finalement :

Lemme 2.4 $\overline{a} = 0$.

Démonstration. Soit $\varepsilon > 0$ et soit n_0 tel que $a_n \leqslant \overline{a} + \varepsilon$ pour $n \geqslant n_0$. Alors pour $n \geqslant n_0$, on a :

$$\begin{aligned}
a_n &\leqslant \zeta_r(N_n^*, X_n^*) + d_n \\
&\leqslant \sum_{k\in[\![0,n_0-1]\!]\cup[\![n-n_0,n-1]\!]} P[Z_n = k] \sup_{\substack{0\leqslant k\leqslant n_0-1 \\ n-n_0\leqslant k\leqslant n-1}} (a_k + a_{n-1-k})\left(\left(\frac{k}{n}\right)^{r/2} + \left(\frac{n-1-k}{n}\right)^{r/2}\right) \\
&\quad + \sum_{k=n_0}^{n-n_0-1} P[Z_n = k]\left[\left(\frac{k}{n}\right)^{r/2}(\overline{a}+\varepsilon) + \left(\frac{n-1-k}{n}\right)^{r/2}(\overline{a}+\varepsilon)\right] + d_n.
\end{aligned}$$

Faisant $n \to \infty$, on obtient :

$$\begin{aligned}
\overline{a} &\leqslant \overline{\lim} \sum_{k\in[\![0,n_0-1]\!]\cup[\![n-n_0,n-1]\!]} P[Z_n = k]\cdot 4a + (\overline{a}+\epsilon)b_r \\
&\leqslant (\overline{a}+\epsilon)b_r,
\end{aligned}$$

puisque la quantité

$$\sum_{k\in[\![0,n_0-1]\!]\cup[\![n-n_0,n-1]\!]} P[Z_n = k] = \frac{2n_0}{n}$$

à n_0 fixé tend vers 0 lorsque n tend vers l'infini. □

Conclusion (fin de la preuve du théorème 2.1). Comme $b_r < 1$, on obtient $\overline{a} = 0$, et la preuve du théorème est complète puisque la ζ_r-convergence entraîne la convergence faible, d'après le théorème B.1de l'annexe B. □

Nous traitons maintenant le cas où la variance est en $n \log n$.

Théorème 2.2 *Si $\alpha = 3/4$, alors X_n converge aussi en distribution vers une loi limite normale.*

Démonstration. On effectue la même preuve que précédemment, avec les mêmes notations, en utilisant

$$\frac{Z_n \log Z_n}{n \log n} \xrightarrow{L_r} \tau \text{ (loi uniforme sur [0,1])}$$

(convergence du moment d'ordre r), ce qui se voit en écrivant

$$\begin{aligned}
\frac{Z_n \log Z_n}{n \log n} &= \frac{Z_n}{n}\frac{\log Z_n}{\log n} \\
&= \frac{Z_n}{n}\frac{1}{\log n}\left(\log n + \log\frac{Z_n}{n}\right) \\
&= \frac{Z_n}{n} + \frac{1}{\log n}\frac{Z_n}{n}\log\frac{Z_n}{n} \\
&= \frac{Z_n}{n} + O\left(\frac{1}{\log n}\right)
\end{aligned}$$

puisque la fonction $x \mapsto x \log x$ est bornée sur $[0,1]$ et que Z_n appartient à $[\![0, n-1]\!]$. □

2.1.3 Cas non gaussiens

Traitons maintenant le cas où la variance n'est pas linéaire ni en $n \log n$, mais en $n^{4\alpha-2}$. Nous montrons que la loi limite n'est pas gaussienne, et nous donnons une équation fonctionnelle caractérisant la fonction caractéristique de la distribution limite.

Théorème 2.3 *Dans le cas $F'(1) > 0$ et $3/4 < \alpha < 1$, $\frac{X_n}{n^{2\alpha-1}}$ tend vers une variable X dont la fonction caractéristique ϕ_X est donnée par*

$$[t]\phi_X(t) = F'(1) \tag{2.5}$$

$$\begin{aligned}
\phi_X(t) &= \alpha\int_0^1 \phi_X(u^{1/\gamma}t)\phi_X((1-u)^{1/\gamma}t)du \\
&\quad +\beta\int_0^1 \phi_X(u^{1/\gamma}t)\phi_X(-(1-u)^{1/\gamma}t)du.
\end{aligned} \tag{2.6}$$

avec $1/\gamma = 1/2 + 2(\alpha - 3/4)$, les relations (2.5) et (2.6) déterminant une unique distribution.

Démonstration. Soit Y une variable aléatoire telle que $\mathrm{P}[Y=1]=\alpha=1-\beta=1-\mathrm{P}[Y=0]$. Alors X_n vérifie l'identité

$$X_n \stackrel{d}{=} Y(A_{Z_n}+B_{n-1-Z_n})+(1-Y)(C_{Z_n}-D_{n-1-Z_n}),$$

avec X,A,B,C,D i.i.d, et $\forall k\in\{0,\dots,n-1\}, P[Z_n=k]=\frac{1}{n}$.

Soit $X_n^* := \frac{X_n}{n^{2\alpha-1}}$ avec $\mathrm{Var}(X_n)\propto n^{4\alpha-2}$.

En normalisant on trouve

$$\begin{aligned} X_n^* \stackrel{d}{=}\ & Y\left(\left(\frac{Z_n}{n}\right)^{1/\gamma}A^*_{Z_n}+\left(\frac{n-1-Z_n}{n}\right)^{1/\gamma}B^*_{n-1-Z_n}\right)\\ & +(1-Y)\left(\left(\frac{Z_n}{n}\right)^{1/\gamma}C^*_{Z_n}-\left(\frac{n-1-Z_n}{n}\right)^{1/\gamma}D^*_{n-1-Z_n}\right). \end{aligned}$$

On peut donc raisonnablement penser qu'on a à la limite

$$\begin{aligned} X \stackrel{d}{=}\ & Y(\tau^{1/\gamma}A+(1-\tau)^{1/\gamma}B)\\ & +(1-Y)(\tau^{1/\gamma}C-(1-\tau)^{1/\gamma}D), \end{aligned}$$

qui en terme de fonction caractéristique se traduit par (2.6).

Nous allons montrer ce résultat en donnant une borne asymptotique sur les moments, dont on donnera une expression récursive en fonction des moments précédents. La méthodologie que nous allons utiliser s'appelle

En dérivant $W_z=\alpha W^2+\beta W\overline{W}$ k fois par rapport à u, on obtient :

$$W_{zu^k}=\sum_{j=0}^{k}\binom{k}{j}W_{u^j}\left(\alpha W_{u^{k-j}}+\beta\overline{W}_{u^{k-j}}\right).$$

Par récurrence, on peut montrer que pour tout k existent des réels e_k et g_k tels que :

$$W_{u^k}=\frac{g_k}{(1-z)^{e_k}}+o\left(\frac{1}{(1-z)^{e_k}}\right),\ g_k>0.$$

De la propriété

$$\exists j<k \text{ tel que } e_k-1=e_j-1+e_{k-j}-1,$$

on montre facilement par récurrence la relation

$$\forall k, e_k=(2\alpha-1)k+1=2\alpha k-(k-1),$$

sachant que les premières valeurs de e_k sont données par : $e_0=1, e_1=2\alpha, e_2=4\alpha-1,\dots$.

Par identification en les termes dominants, on a la relation de récurrence

$$e_k g_k=\sum_{j=0}^{k}\binom{k}{j}g_j\left(\alpha g_{k-j}+\beta(-1)^{k-j}g_{k-j}\right), \tag{2.7}$$

avec $g_0=1$ et $g_1=f'$.

Soit m_k la limite du k^e moment de la variable X_n^*. Le moment d'ordre j de X_n^* étant négligeable devant celui d'ordre k pour tout $j<k$, on a

$$m_k=\lim_{n\to\infty}\frac{[z^n]W_{u^k}(z,1)}{n^{k(2\alpha-1)}}=\frac{g_k}{\Gamma((2\alpha-1)k+1)}. \tag{2.8}$$

En écrivant

$$\phi_X(t) = \sum_k m_k \frac{(-t)^k}{k!},$$

et en utilisant (2.7) et la relation (propriété des fonctions bêta)

$$\forall k,\ \frac{1}{\Gamma\left(\frac{n}{\gamma}+1\right)} = \int_0^1 \frac{u^{k/\gamma}}{\Gamma\left(\frac{k}{\gamma}+1\right)} \frac{u^{(n-k)/\gamma}}{\Gamma\left(\frac{n-k}{\gamma}+1\right)}\, du,$$

on obtient (2.6), sous réserve que $\lim X_n^*$ existe.

Pour montrer cette dernière condition, il suffit ([63] vol. 2 p.514) de montrer

$$\overline{\lim}\frac{1}{n} m_n^{1/n} < \infty.$$

Pour ce faire, introduisons la suite (h_i) définie par $h_0 = g_0$, $h_1 = g_1$ et

$$e_2 h_k = \sum_{j=0}^{k} \binom{k}{j} h_j \left(\alpha h_{k-j} + \beta \left|(-1)^{k-j}\right| h_{k-j}\right) = \sum_{j=0}^{k} \binom{k}{j} h_j h_{k-j}. \tag{2.9}$$

On montre facilement par récurrence, en utilisant la propriété $e_2 \leqslant e_k$ pour $k \geqslant 2$, que (h_i) est une suite majorante de $(|g_i|)$, si bien que la série H, définie par

$$H(x) = \sum_n \frac{h_n}{n!} x^n,$$

vérifie

$$\begin{aligned} e_2 H(x) &= H^2 + e_2(h_0 + h_1 x) - h_0^2 - 2h_0 h_1 x \\ H(x) &= \frac{e_2 - \sqrt{e_2^2 - 4\left(e_2(h_0 + h_1 x) - h_0^2 - 2h_0 h_1 x\right)}}{2} \\ &= \frac{e_2 - \sqrt{(e_2 - 2h_0)^2 - 4h_1(e_2 - 2h_0)x}}{2}. \end{aligned}$$

Notons que $e_2 > 2$ équivaut à $\alpha > 3/4$, si bien que l'expression de H ci-dessus est bien valide, avec en particulier $H(0) = 1$.

Si on définit η par

$$\eta = \frac{e_2 - 2h_0}{4h_1},$$

qui est strictement positif comme on peut le voir en écrivant

$$\eta = \frac{e_2 - 2}{|f'|} = \frac{4\alpha - 3}{|f'|},$$

alors, par analyse de singularité appliquée à H,

$$\frac{h_n}{n!} \sim \frac{\eta^n}{\sqrt{\pi n^3}}.$$

On peut donc borner les moments m_n définis plus haut

$$|m_n| = \frac{|g_n|}{\Gamma((2\alpha-1)n+1)} \leqslant \frac{n!}{\Gamma((2\alpha-1)n+1)}\frac{h_n}{n!} \sim \frac{n!}{\Gamma((2\alpha-1)n+1)}\frac{\eta^n}{\sqrt{\pi n^3}}$$
$$\sim \frac{\left(\frac{n}{e}\right)^n\sqrt{2\pi n}}{\left(\frac{(2\alpha-1)n+1}{e}\right)^n\sqrt{2\pi\left((2\alpha-1)n+1\right)}}\frac{\eta^n}{\sqrt{\pi n^3}},$$

d'où

$$|m_n|^{1/n} = \left(\frac{|g_n|}{\Gamma((2\alpha-1)n+1)}\right)^{1/n} \leqslant \left(\frac{n!}{\Gamma((2\alpha-1)n+1)}\frac{h_n}{n!}\right)^{1/n} \sim \frac{n}{e}\frac{e}{(2\alpha-1)n+1}\eta$$
$$\sim \frac{\eta}{2\alpha-1}.$$

Donc $\lim \frac{1}{n}|m_n|^{1/n} = 0$, et on a convergence des moments de tous les ordres de X_n^* vers X définie par (2.5) et (2.6). □

Remarque 2.1 *On peut remarquer que les lois stables de fonction caractéristique $e^{-b|t|^\gamma}$ vérifient (2.6) mais ne vérifient pas (2.5), et ne sont donc pas solutions. En outre, les lois stables ne possèdent en général pas de moment fini de tout ordre, ce qui les écarte d'autant plus.*

Remarque 2.2 *On peut caractériser X directement par (2.7) à la place de (2.6), ce qui sur la série $G(x) = \sum_n \frac{g_n}{n!}(-x)^n$, se traduit par*

$$-(2\alpha-1)xG'(x) + G(x) = \alpha G^2(x) + \beta G(x)G(-x),$$

et, en utilisant (2.8) et $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$,

$$G(x) = \int_0^\infty e^{-t}\phi_X(xt^{2\alpha-1})\,dt = \int_0^\infty e^{-(xu)^{1/(2\alpha-1)}}x^{1/(2\alpha-1)}u^{1/(2\alpha-1)-1}\phi_X(u)\,du.$$

On peut noter que la distribution limite n'est pas symétrique :

Corollaire 2.1 *La variable aléatoire X définie dans le théorème 2.3 suit une distribution asymétrique.*

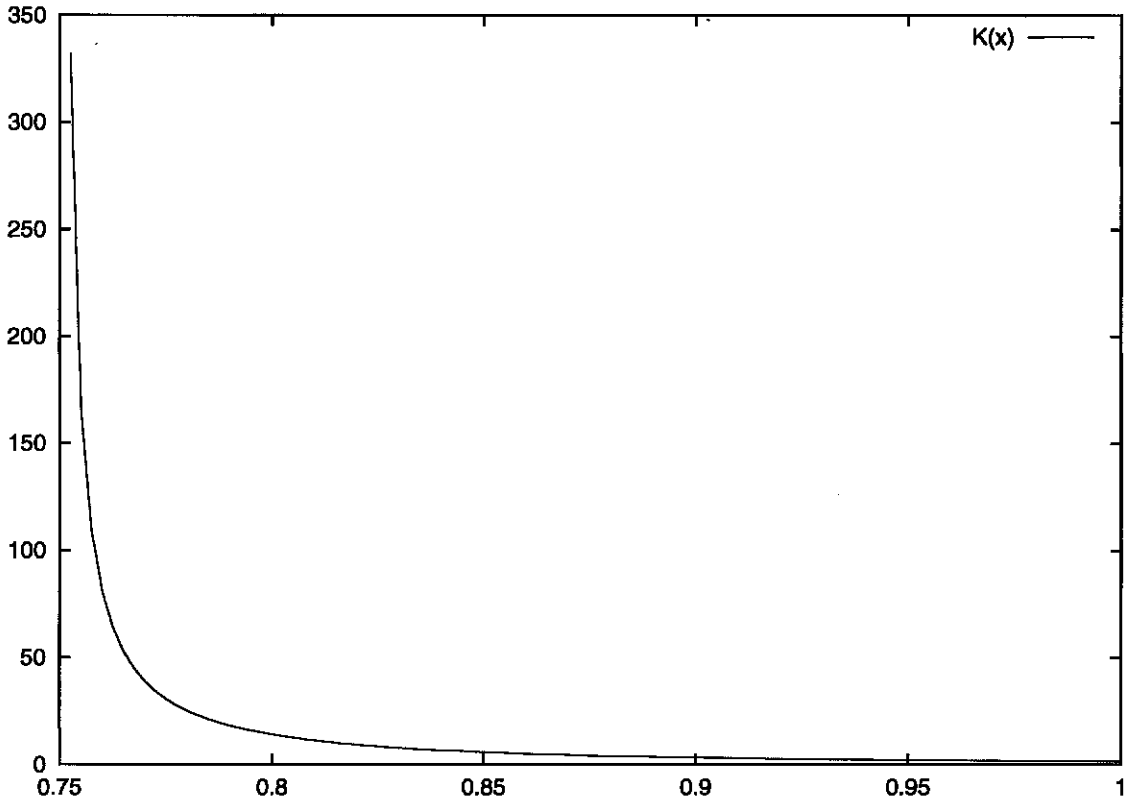
Démonstration. Avec les notations précédentes, on trouve

$$g_2 = C_\alpha f'^2 = \frac{\alpha-1/2}{\alpha-3/4}f'^2,$$
$$g_3 = \frac{3.2\alpha g_1 g_2}{(2\alpha-1)\cdot 3 - 2\alpha} = \frac{6\alpha}{4\alpha-3}g_1g_2 = \frac{6\alpha}{4\alpha-3}C_\alpha f'^3,$$

si bien que le cumulant d'ordre 3 vérifie

$$\kappa_3 = \mathbf{E}[(X-m_1)^3]$$
$$= m_3 - 3m_1m_2 + 2m_1^3$$
$$= \frac{g_3}{\Gamma(6\alpha-2)} - 3\frac{g_1}{\Gamma(2\alpha)}\frac{g_2}{\Gamma(4\alpha-1)} + 2\frac{g_1^3}{\Gamma(2\alpha)^3}$$
$$= \frac{6\alpha}{4\alpha-3}\frac{C_\alpha f'^3}{\Gamma(6\alpha-2)} - 3\frac{C_\alpha f'^3}{\Gamma(2\alpha)\Gamma(4\alpha-1)} + \frac{2f'^3}{\Gamma(2\alpha)^3}.$$

Or la fonction $K : \alpha \mapsto \kappa_3 f'^{-3}$ est non nulle sur $]3/4, 1]$. Voir la courbe



Le moment d'ordre 3 de la variable $X - \mathbf{E}[X]$ est donc non nulle, et il en résulte que sa distribution est asymétrique. □

2.1.4 Perspectives : analyse de singularité

La question qui se pose est de savoir si on peut retrouver les résultats ci-dessus par analyse de singularités. Pour $\alpha = 0$, on peut retrouver une loi limite normale par le théorème des quasi-puissances.

Théorème 2.4 *Si $\alpha = 0$, la valuation des ABR avec des + et des − suit une loi limite gaussienne.*

Démonstration. On a l'équation fonctionnelle

$$W_z = W\overline{W},$$

de condition initiale

$$W(0, u) = F.$$

Par symétrie, on a $W - F = \overline{W} - \overline{F}$, d'où l'équation de Ricatti

$$W_z = W(W - F + \overline{F}),$$

de solution :

$$\frac{1}{W(z,u)} = \frac{1}{F} e^{(F-\overline{F})z} + \frac{1}{F - \overline{F}} \left(1 - e^{(F-\overline{F})z}\right).$$

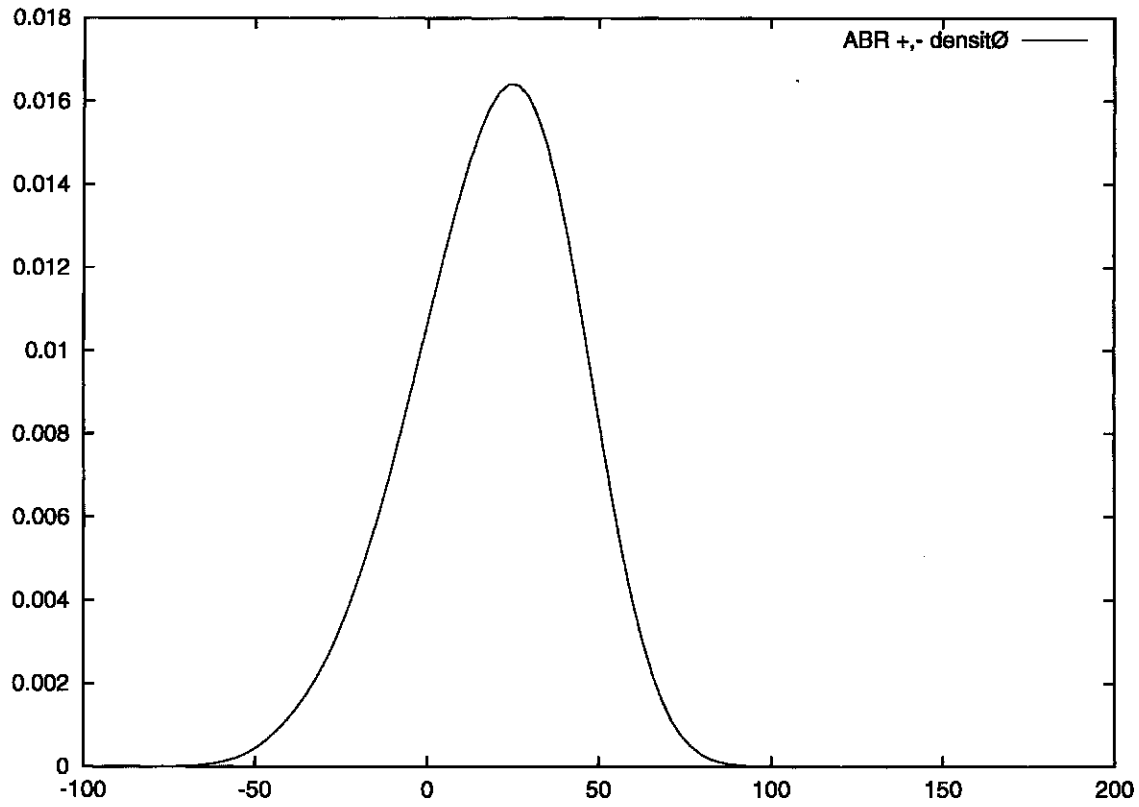


FIG. 2.1 – Fonction densité pour $\alpha = 0,85$ et $F(x) = x$.

La singularité dominante est donnée par la singularité simple

$$z = \rho(u) = \frac{1}{F - \overline{F}} \log\left(\frac{F}{\overline{F}}\right),$$

avec $\rho(1) = 1$, $\rho'(1) = 0$ et $\rho''(1) = \mathrm{Var}(F) + \frac{1}{3}F'(1)^2$, si bien que la condition de variabilité est vérifiée. On a la limite gaussienne en appliquant le théorème des quasi-puissances (rappelé dans l'annexe A). □

Il apparaît encore difficile de traiter le cas $\alpha \in]0,1[$ par séries génératrices et analyse de singularité uniquement.

2.2 Conclusion

Avec des opérateurs "+" et "−", si on tire la structure de l'arbre de manière non uniforme mais selon la distribution des arbres de recherche, on obtient dans la plupart des cas une limite gaussienne, mais on obtient aussi parfois une loi limite exotique, dans le cas où la probabilité de l'opérande "+" est entre 3/4 et 1 et où la moyenne des variables aléatoires placées aux feuilles est non nulle.

Il serait intéressant de pouvoir donner une interprétation physique de l'apparition d'une telle loi, et en particulier de trouver le pourquoi d'un phénomène de seuil en $\alpha = 3/4$.

Une question d'ordre plus technique consisterait à pouvoir retrouver ces résultats en utilisant seulement des outils de variable complexe, sans avoir recours aux méthodes de point fixe.

Dans le chapitre suivant, nous démontrerons à nouveau le caractère gaussien de la limite dans le cas $\alpha < 1/2$, en utilisant des méthodes de martingales.

Chapitre 3

Martingales

Dans les chapitres précédents, nous avons étudié des paramètres définis de manière récursive sur les arbres, ce qui se prêtait bien à l'utilisation de séries génératrices ou de méthodes de point fixe : le caractère récursif des paramètres se retrouve dans le produit de séries, ou encore dans des équations limites portant sur des distributions de probabilités.

Certains de ces paramètres peuvent être vus d'une autre manière, à savoir d'une manière itérative. Pour ces paramètres, il est possible, à partir d'un arbre de taille n, de définir un algorithme permettant de transformer cet arbre en un arbre de taille $n+1$, et de décrire comment le paramètre considéré varie au cours de cette transformation.

Pour ces paramètres, il est alors possible d'étudier la distribution limite à l'aide de martingales, les arbres de taille n formant alors une filtration au temps n. Par manque de place il ne nous est pas possible d'effectuer des rappels sur les martingales. Le lecteur non familier avec les martingales pourra se référer aux livres d'introduction [7] ou [210], ainsi qu'au cours [29].

Dans ce chapitre, nous utilisons le théorème central limite par martingales [53, 54, 60, 104] afin de donner une autre preuve du caractère gaussien d'une part de la distribution limite des nombres de Narayana, et d'autre part de la distribution limite de la valuation des ABR avec des "+" placés avec probabilité $\alpha < \frac{1}{2}$. Le seuil $\alpha = \frac{1}{2}$ provient de ce que, comme on le verra plus bas, on peut démontrer une condition de Lindeberg pour $\alpha < \frac{1}{2}$.

3.1 Rappels sur le théorème central limite par martingales

On rappelle les théorème 3.2 et corollaire 3.1 de Hall-Heyde [104] p.58-59 :

Théorème 3.1 (**3.2** de Hall-Heyde) *Soit* $\{S_{ni},\ \mathcal{F}_{ni},\ 1 \leqslant i \leqslant k_n,\ n \geqslant 1\}$ *un tableau de martingale de moyenne nulle et de carré intégrable, aux différences* X_{ni}, *et soit* μ^2 *une variable aléatoire presque sûrement finie. Supposons que*

$$\max_i |X_{ni}| \xrightarrow{P} 0, \tag{3.18}$$

$$\sum_i X_{ni}^2 \xrightarrow{P} \mu^2, \tag{3.19}$$

$$\mathbf{E}\left(\max_i X_{ni}^2\right) \text{ est bornée en } n, \tag{3.20}$$

et que les $\sigma-$*filtrations sont emboîtés :* $\mathcal{F}_{ni} \subseteq \mathcal{F}_{n+1,i}$ *pour* $1 \leqslant i \leqslant k_n,\ n \geqslant 1$. *(3.21)*
Alors $S_{nk_n} = \sum_i X_{ni} \xrightarrow{d} Z$, *où la variable aléatoire* Z *a pour fonction caractéristique* $\mathbf{E}\left[e^{-\frac{1}{2}\mu^2 t^2}\right]$.

Le lecteur trouvera une définition des tableaux de martingales dans le même livre [104], p. 52-53.

La notation $\xrightarrow{P}$ indique la convergence en probabilité [23, 91].

Dans les cas qui nous intéressent, la variable aléatoire μ sera en fait constante.

Corollaire 3.1 (**3.1** de Hall-Heyde) *Si (3.18) et (3.20) sont remplacés par la condition de Lindeberg conditionelle (3.7) :*

$$\textit{Pour tout } \varepsilon > 0, \qquad \sum_i \mathbf{E}\left[X_{ni}^2 I\left(|X_{ni}| > \varepsilon\right) | \mathcal{F}_{n,i-1}\right] \xrightarrow{P} 0,$$

si (3.19) est remplacé par une condition analogue sur la variance conditionnelle :

$$V_{nk_n}^2 = \sum \mathbf{E}\left(X_{ni}^2 | \mathcal{F}_{n,i-1}\right) \xrightarrow{P} \mu^2$$

et si (3.21) est vérifié, alors la conclusion du théorème 3.2 reste vraie.

3.2 Distribution des nombres de Narayana

Dans cette section, nous retrouvons la nature de la distribution limite des nombres de Narayana par argument de martingales. Nous considérons les nombres de Narayana comme paramètres sur des arbres de Catalan. Nous passons d'un arbre de Catalan de taille n à un arbre de Catalan de taille $n+1$ par l'algorithme de Rémy, et nous verrons que les nombres de Narayana se comportent de manière sympathique par rapport à ce processus de croissance des arbres.

3.2.1 Algorithme de Rémy

Pour engendrer un arbre de Catalan de manière uniforme, il existe l'algorithme de Rémy, décrit dans [182], qui consiste à faire pousser un arbre de la manière suivante : on choisit un nœud a, interne ou externe, au hasard dans l'arbre de Catalan T_n de taille n. Soit S le sous-arbre de racine a. On le remplace de manière équiprobable par le sous-arbre $\square \cdot S$ ou le sous-arbre $S \cdot \square$ pour obtenir un arbre T_{n+1} de taille $n+1$. Cette procédure se traduit par la formule :

$$C_{n+1} = \frac{2(2n+1)}{n+2} C_n.$$

Le 2 indique les deux possibilités d'ajouter une feuille gauche ou une feuille droite, $(2n+1)$ énumère les nœuds internes et externes de l'arbre T_n, et $(n+2)$ pointe sur la dernière feuille ajoutée dans l'arbre T_{n+1}.

3.2.2 Les nombres de Narayana comme martingales

L'algorithme ci-dessus permet de passer d'un arbre de taille n à un arbre de taille $n+1$. On peut considérer l'ensemble des arbres de taille n comme une filtration à un instant n. Nous allons montrer que les nombres de Narayana possèdent la propriété de martingales suivant cette filtration.

Théorème 3.2 *Soit L_n le nombre de feuilles gauches d'un arbre de Catalan taille n. Soient les suites (P_n) et (a_n) définies par*

$$P_n = \prod_{j=1}^{n} \left(1 - \frac{1}{2j+1}\right),$$

et

$$\begin{aligned} a_0 &= 1, \\ a_{n+1} &= \left(1 - \frac{1}{2(2n+1)}\right) a_n + \frac{1}{2} \textit{ pour } n \geqslant 0. \end{aligned}$$

Alors la variable $S_n := \dfrac{L_n}{P_n} - a_n$ *est une martingale.*

Démonstration. Supposons qu'on passe d'un arbre de taille n à un arbre de taille $n+1$ en suivant l'algorithme de Rémy. Soient AG l'événement "l'ancien nœud est une feuille gauche", AD l'événement "l'ancien nœud est une feuille droite", AT l'événement "l'ancien nœud est la racine d'un sous-arbre de taille non nulle", NG l'événement "la nouvelle feuille est une feuille gauche", ND l'événement "la nouvelle feuille est une feuille droite".

On peut comparer ce phénomène à un modèle d'urne : un arbre contient un certain nombre de types de feuille. On peut donc l'assimiler à une urne contenant des boules de couleurs différentes. Suivant la couleur de la boule qu'on tire, on va rajouter une boule d'une certaine couleur.

Le tableau qui donne la variation de L_n suivant l'opération effectuée s'écrit :

	AG	AD	AT
NG	0	+1	+1
ND	0	+1	0

Chaque case du tableau correspond à la conjonction de deux événements, et le contenu de la case représente la variation de L_n. Par exemple, dans la case (NG,AD), qui signifie que l'ancien nœud est une feuille droite (événement AD), et qu'on produit une nouvelle feuille gauche, on a un contenu "+1" car le nouvel arbre contient une feuille gauche de plus que l'ancien arbre. Notons qu'on a aussi un "+1" dans la case (ND,AD) car dans ce cas, même si la nouvelle feuille qu'on ajoute est une feuille droite, l'ancienne feuille droite devient une feuille gauche et le nouvel arbre contient aussi une feuille gauche de plus que l'ancien arbre.

Rappelons qu'un arbre de taille n possède $2n+1$ nœuds dont n internes et $n+1$ externes, ces derniers se répartissant en L_n feuilles gauches et $n+1-L_n$ feuilles droites. Les probabilités de chacun de ces petits événements sont donc données par

$$P_{\mathrm{NG}} = P_{\mathrm{ND}} = \frac{1}{2}, \quad P_{\mathrm{AG}} = \frac{L_n}{2n+1}, \quad P_{\mathrm{AD}} = \frac{n+1-L_n}{2n+1}, \quad P_{\mathrm{AT}} = \frac{n}{2n+1}.$$

Le nombre L_{n+1} de feuilles gauches d'un arbre T_{n+1} se déduit de l'arbre T_n à L_n feuilles gauches par la relation

$$L_{n+1} = L_n + \mathbb{I}_{\mathrm{NG}}\mathbb{I}_{\neg\mathrm{AG}} + \mathbb{I}_{\mathrm{ND}}\mathbb{I}_{\mathrm{AD}},$$

d'où :

$$\begin{aligned} \mathrm{E}[L_{n+1}|T_n] &= L_n + \frac{1}{2}\left(1 - \frac{L_n}{2n+1}\right) + \frac{1}{2}\frac{n+1-L_n}{2n+1} \\ &= \left(1 - \frac{1}{2n+1}\right) L_n + \frac{3n+2}{4n+2}. \end{aligned}$$

Posant $P_n = \displaystyle\prod_{j=1}^{n}\left(1 - \frac{1}{2j+1}\right)$, on obtient :

$$\mathrm{E}\left[\frac{L_{n+1}}{P_{n+1}}\middle| T_n\right] = \frac{L_n}{P_n} + \frac{1}{2P_{n+1}}$$

avec $L_0 = L_1 = 1$.

Si on définit (a_n) par $a_0 := 1$ et $a_{n+1} := \left(1 - \dfrac{1}{2(2n+1)}\right) a_n + \dfrac{1}{2}$ pour $n \geqslant 0$, la variable $S_n := \dfrac{L_n}{P_n} - a_n$ est donc une martingale, et $X_n := S_n - S_{n-1}$, $n \geqslant 1$ une martingale aux différences. □

3.2.3 Distribution limite des nombres de Narayana par les martingales

On utilise la méthodologie développée par [145, 148, 150, 149, 147, 191].

Théorème 3.3 *Les nombres de Narayana, après normalisation, convergent en distribution vers une loi gaussienne.*

Démonstration. En utilisant les notations précédentes, on a

$$\mathrm{E}[L_{n+1}|T_n] = \left(1-\frac{1}{2n+1}\right)L_n + \frac{3n+2}{4n+2}.$$

Posons $X_n = L_n - \mu_n = L_n - \frac{n+1}{2}$, qui vérifie donc

$$\mathrm{E}[X_{n+1}|T_n] = \left(1-\frac{1}{2n+1}\right)X_n$$

Notons qu'on peut retrouver μ_n à partir de la relation de récurrence

$$\mu_{n+1} = \left(1-\frac{1}{2n+1}\right)\mu_n + \frac{3n+2}{4n+2}$$

et des conditions initiales $\mu_0 = \mu_1 = 0$ (la valeur de μ_0 est indifférente).

En effet, si on pose $g(x) = \sum \mu_n x^n$, il vient

$$\begin{aligned}
(2n+1)\mu_{n+1} &= 2n\mu_n + \frac{3n+2}{2} \\
2(n+1)\mu_{n+1} - \mu_{n+1} &= 2n\mu_n + \frac{3n+2}{2} \\
2g'(x) - x^{-1}(g(x)-1) &= 2xg'(x) + \frac{3x}{2(1-x)^2} + \frac{1}{1-x}
\end{aligned}$$

de solution générale, lorsque l'expression ci-dessous est définie :

$$g(x) = 1/2\,\frac{x^2}{(-1+x)^2} - \frac{x}{(-1+x)^2} + (-1+x)^{-2} + C\frac{\sqrt{x}}{\sqrt{-1+x}},$$

avec C une constante. On veut que g soit analytique donc la solution cherchée est :

$$g(x) = 1/2\,\frac{x^2}{(1-x)^2} - \frac{x}{(1-x)^2} + (1-x)^{-2},$$

de terme général $g_n = \mu_n = \frac{n+1}{2}$ pour $n \geq 1$.

De

$$\mathrm{E}[X_{n+1}|T_n] = \left(1-\frac{1}{2n+1}\right)X_n,$$

on tire

$$\mathrm{E}[X_n|T_{n-1}] = \left(1-\frac{1}{2n-1}\right)X_{n-1},$$

puis

$$\mathrm{E}[X_n - X_{n-1}|T_{n-1}] = \left(-\frac{1}{2n-1}\right)X_{n-1}.$$

Soit (ΔM_n) définie par

$$\begin{aligned}\Delta M_n &= \Delta X_n - E[\Delta X_n | T_{n-1}] \\ &= X_n - X_{n-1} - E[X_n - X_{n-1} | T_{n-1}] \\ &= X_n - X_{n-1} - \frac{1}{2n-1} X_{n-1} \\ &= X_n - \left(\frac{2n-1}{2n-1} - \frac{1}{2n-1}\right) X_{n-1} \\ &= X_n - X_{n-1}\left(\frac{2n-2}{2n-1}\right).\end{aligned}$$

Soient pour tout n la suite $(b_{i,n})_{1 \leqslant i \leqslant n}$ et le nombre c_n tels que

$$\begin{aligned}\sum_{i=1}^{n} b_{in}\left[X_i - X_{i-1}\left(\frac{2i-2}{2i-1}\right)\right] &= X_n + c_n, \\ b_{n,n}X_n + \sum_{i=1}^{n-1}\left[b_{i,n} - b_{i+1,n}\frac{2i}{2i+1}\right] X_i - b_{1,n}\frac{2\cdot 1 - 2}{2\cdot 1 - 1} &= X_n + c_n,\end{aligned}$$

(après une transformation d'Abel), d'où : $c_n = 0$, $b_{n,n} = 1$ et $\forall i \in [\![1, n-1]\!]$, $b_{i,n} = b_{i+1,n}\dfrac{2i}{2i+1}$.
On a

$$\frac{\max b_{i,n}^2}{\sum_{i=1}^n b_{i,n}^2} \xrightarrow[n\to\infty]{} 0 \quad \left(, \text{ qui entraîne aussi } \frac{\max b_{in}}{\sqrt{\sum_{i=1}^n b_{in}^2}} \xrightarrow[n\to\infty]{} 0\right).$$

En effet, soit $k \in \mathbb{N}^*$. On a, pour tout $i < k$,

$$\lim_{n\to\infty} b_{i,n} = b_{n,n} = 1,$$

donc

$$\lim_{n\to\infty} \sum_{i=0}^{k-1} b_{n-i}^2 = k$$

et

$$\overline{\lim_n} \frac{\max b_{in}^2}{\sum_{i=1}^n b_{i,n}^2} \leq \frac{1}{k}.$$

Appliquons le théorème de Hall-Heyde avec $X_{ni} = \dfrac{b_{in}\Delta M_i}{\sqrt{\sum_{i=1}^n b_{i,n}^2}}$.

Les ΔM_i étant bornés, on a aussi $\max_i \dfrac{b_{in}|\Delta M_i|}{\sqrt{\sum_{i=1}^n b_{in}^2}} \xrightarrow[n\to\infty]{P} 0$, ainsi que le fait que

$$\mathrm{E}\left[\max_i \frac{b_{i,n}^2(\Delta M_i)^2}{\sum_{i=1}^n b_{i,n}^2}\right]$$

est uniformément borné en n.

$$\begin{aligned}
\mathrm{E}\left[(\Delta X_n)^2|T_{n-1}\right] &= \mathrm{E}\left[(X_n - X_{n-1})^2|T_{n-1}\right] \\
&= \mathrm{E}\left[(L_n - \mu_n - (L_{n-1} - \mu_{n-1}))^2|T_{n-1}\right] \\
&= \mathrm{E}\left[(L_n - L_{n-1})^2|\, T_{n-1}\right] - 2\mathrm{E}\left[(L_n - L_{n-1})(\mu_n - \mu_{n-1})|T_{n-1}\right] + (\mu_n - \mu_{n-1})^2 \\
&= \mathrm{E}\left[L_n - L_{n-1}|T_{n-1}\right] - 2\frac{1}{2}\mathrm{E}\left[L_n - L_{n-1}|T_{n-1}\right] + \frac{1}{4} \text{ (car } L_n - L_{n-1} \in \{0,1\}) \\
&= \frac{1}{4}.
\end{aligned}$$

On a

$$\begin{aligned}
\mathrm{E}\left[(\Delta M_n)^2|T_{n-1}\right] &= \mathrm{E}\left[(\Delta X_n)^2|T_{n-1}\right] - \mathrm{E}\left[\Delta X_n|T_{n-1}\right]^2 \\
&= \mathrm{E}\left[(\Delta X_n)^2|T_{n-1}\right] - \left(\frac{X_{n-1}}{2n-1}\right)^2 \\
&\xrightarrow[n\to\infty]{moy} \frac{1}{4}, \text{ car la variance } \mathrm{E}[X_{n-1}^2] \text{ est linéaire.}
\end{aligned}$$

Par le théorème de Césaro, on a aussi

$$\sum_i \frac{b_{in}^2(\Delta M_i)^2}{\sum_i b_{in}^2} \xrightarrow[n\to\infty]{moy} \frac{1}{4}.$$

et à plus forte raison

$$\sum_i \frac{b_{in}^2(\Delta M_i)^2}{\sum_i b_{in}^2} \xrightarrow[n\to\infty]{P} \frac{1}{4}.$$

Donc

$$\sum_i \frac{b_{in}\Delta M_i}{\sqrt{\sum_{i=1}^n b_{in}^2}} = \frac{X_n}{\sqrt{\sum_{i=1}^n b_{in}^2}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \frac{1}{4}).$$

En regardant le moment d'ordre 2 on a :

$$\mathrm{E}\left[\frac{X_n^2}{\sum_{i=1}^n b_{in}^2}\right] \xrightarrow[n\to\infty]{} \frac{1}{4}.$$

Or on sait que $\mathrm{E}\left[X_n^2\right] = \sigma_n^2 \sim 4\frac{n}{8} \sim \frac{n}{2}$, donc $\sum_{i=1}^n b_{in}^2 \sim \frac{n}{2}$ et $\sqrt{\sum_{i=1}^n b_{in}^2} \sim \sqrt{\frac{n}{2}}$. Finalement,

$$\frac{X_n}{\sqrt{n/2}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \frac{1}{4}),$$

et

$$\frac{X_n}{\sqrt{n/8}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

□

3.3 ABR +, − et martingales

Dans ce paragraphe, nous étudions la distribution limite des résultats d'expressions arithmétiques formées sur des arbres binaires de recherche avec opérateurs + et −, lorsque l'opérateur + apparaît avec une probabilité α strictement inférieure à 1/2. Nous montrons en utilisant le théorème central limite sur les martingales, qu'à cette dernière condition, nous obtenons une loi limite gaussienne. La technologie du théorème central limite sur les martingales apparaît donc *a priori* moins puissante que celle du théorème du point fixe.

La condition $\alpha < 1/2$ apparaît de manière cruciale lorsqu'on veut que la condition de Lindeberg soit réalisée. Remarquons également que la moyenne de la valuation, en $n^{2\alpha-1}$, est alors évanescente.

Soit Y_n la valuation de l'arbre de taille n, avec Y_0 une variable aléatoire de fonction génératrice F analytique en 1.

On note $S(n)$ la variable aléatoire notant le nombre de nœuds externes remontant positivement à la racine (c'est-à-dire que la variable aléatoire Y_0 se trouvant à ce nœud externe garde le même signe dans l'évaluation de l'expression arithmétique finale), et $\overline{S}(n) = n + 1 - S(n)$ le nombre de nœuds externes remontant négativement à la racine. Autrement dit, pour tout n, la variable Y_n a la distribution :

$$Y_n \stackrel{d}{=} \sum_{k=1}^{S(n)} Y_{0k} - \sum_{k=1}^{\overline{S}(n)} Y'_{0k},$$

où les variables Y_{0k} et Y'_{0k} sont des variables indépendantes de même distribution que Y_0.

3.3.1 Moments de Y_n et S_n

Nous verrons que les incréments de Y_n dépendent de ceux de S_n, si bien qu'il est utile de connaître les moments joints d'ordre 1 et 2 de Y_n et S_n.

Proposition 3.1 *Notons* $f' = F'(1)$, $f'' = F''(1)$, $C_\alpha = \frac{\alpha-\frac{1}{2}}{\alpha-\frac{3}{4}}$ *et* $\lambda_0 = f'' - C_\alpha f'^2 + f'$. *Alors asymptotiquement, pour* $r+s \leqslant 2$, *le terme dominant du moment joint* $\mathbf{E}[Y_n^r S_n^s]$ *de* Y_n *et* S_n *est donné par la case située à la ligne* r *et la colonne* s *du tableau*

$Y_n \backslash S_n$	0	1	2
0	1	$\frac{n}{2}$	$\frac{1}{4}\lambda_0 n$
1	$f'\frac{n^{2\alpha-1}}{\Gamma(2\alpha)}$	$\frac{1-C_\alpha}{2}n$	
2	$\lambda_0 n$		

Démonstration. Nous connaissons déjà les moments de Y_n par les calculs effectués dans le chapitre 3. En remarquant que pour $Y_0 \equiv 1$, nous avons $Y_n = S(n) - \overline{S}(n) = 2S_n - (n+1)$, nous

avons directement la moyenne et la variance de S_n :

$$\begin{aligned}
\mathbf{E}[S_n] &= \mathbf{E}\left[\frac{Y_n+n+1}{2}\right] = \mathbf{E}[Y_n] + \frac{n+1}{2} = \frac{\Gamma(n+2\alpha)}{\Gamma(2\alpha)n!} \\
&\underset{n\to\infty}{\sim} \frac{n}{2}, \\
\mathbf{E}\left[S_n^2\right] &= \mathbf{E}\left[\frac{Y_n^2}{4} + \frac{Y_n(n+1)}{2} + \frac{(n+1)^2}{4}\right] \\
&= \frac{1}{4}\mathbf{E}\left[Y_n^2\right] + \frac{n+1}{2}\mathbf{E}[Y_n] + \frac{(n+1)^2}{4}, \\
\mathrm{Var}[S_n] &\underset{n\to\infty}{\sim} \frac{1}{4}\lambda_0 n + \frac{(n+1)^2}{4} - \frac{(n+1)^2}{4} \\
&\underset{n\to\infty}{\sim} \frac{1}{4}\lambda_0 n.
\end{aligned}$$

Pour calculer la covariance de Y_n et S_n, il apparaît judicieux d'utiliser des séries génératrices, ce qui au passage nous fera retrouver les moyennes de Y_n et S_n. On introduit $W(z,u,v)$, la série génératrice des arbres binaires de recherche, où z énumère la taille, u la valuation Y_n, et v le nombre S_n de nœuds externes remontant positivement. En dérivant

$$W_z(z,u,v) = \alpha W(z,u,v)^2 + \beta v W(z,u,v) W(zv, u^{-1}, v^{-1})$$

le nombre de fois idoine par rapport à u et v, et en faisant $u = v = 1$, nous obtenons des équations différentielles linéaires qui se résolvent en

$$\begin{aligned}
W_u &= \frac{f'}{(1-z)^{2\alpha}}, \\
W_v &= \frac{1}{2}\left[\frac{1}{(1-z)^{2\alpha}} + \frac{1}{(1-z)^2}\right], \\
W_{uv} &= \frac{2\alpha-1}{4\alpha-3}\frac{f'}{(1-z)^{4\alpha-1}} + \frac{\alpha f'}{(1-z)^{2\alpha+1}} + \left(\frac{1}{2}-\alpha\right)\frac{f'}{(1-z)^{2\alpha}} + \frac{1}{2}\frac{1-C_\alpha}{(1-z)^2}.
\end{aligned}$$

□

Lemme 3.1 *Soit T_n la filtration des arbres de taille n. Nous avons les relations :*

$$\mathbf{E}[S_{n+1}|T_n] = \left(1 + \frac{2\alpha-1}{n+1}\right) S_n + 1 - \alpha, \tag{3.1}$$

$$\mathbf{E}[Y_{n+1}|T_n] = Y_n + f'(2\alpha-1)\left(\frac{2S_n}{n+1} - 1\right). \tag{3.2}$$

Démonstration. Pour passer d'un arbre de taille n à un arbre de taille $n+1$, on choisit un nœud externe au hasard, de manière uniforme, parmi $n+1$. Dans l'arbre de taille n, ce nœud remonte positivement à la racine, *i.e.* sa valeur est comptée positivement dans le calcul final de la valuation (événement $\mathbb{1}_>$, contraire de $\mathbb{1}_<$) avec la probabilité $\frac{S_n}{n+1}$. Dans l'arbre de taille $n+1$, il est remplacé par nœud interne comportant un opérateur $+$ avec la probabilité α (événement

$\mathbb{1}_+$, d'événement contraire $\mathbb{1}_-$). Ceci s'écrit

$$\begin{aligned}
S_{n+1} &= S_n + \mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_< \\
\mathbf{E}[S_{n+1}|T_n] &= S_n + \alpha\frac{S_n}{n+1} + \beta\left(1 - \frac{S_n}{n+1}\right) \\
&= \left(1 + \frac{2\alpha - 1}{n+1}\right) S_n + 1 - \alpha \\
Y_{n+1} &= Y_n + Y_0\left(\mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_< - \mathbb{1}_+\mathbb{1}_< + \mathbb{1}_-\mathbb{1}_>\right) \\
&= Y_n + Y_0\left(2\left(\mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_<\right) - 1\right) \\
\mathbf{E}[Y_{n+1}|T_n] &= Y_n + f'(2\alpha - 1)\left(\frac{2S_n}{n+1} - 1\right)
\end{aligned}$$

□

Théorème 3.4 *Y_n possède une distribution limite normale.*

Démonstration. Nous allons en fait montrer que (Y_n, S_n) suit une loi limite normale, pour en déduire que la marginale Y_n est également normale. On introduit les notations $y_n = \mathbf{E}[Y_n]$, $s_n = \mathbf{E}[S_n]$, $\Delta Y_n = Y_n - Y_{n-1}$, $\Delta S_n = S_n - S_{n-1}$, $X_n = Y_n - y_n$, $Z_n = S_n - s_n$, $\Delta M_i^X = \Delta X_i - \mathbf{E}[\Delta X_i|T_{i-1}]$ et $\Delta M_i^Z = \Delta Z_i - \mathbf{E}[\Delta Z_i|T_{i-1}]$. De

$$\begin{aligned}
\mathbf{E}[S_{n+1}|T_n] &= \left(1 + \frac{2\alpha - 1}{n+1}\right) S_n + 1 - \alpha, \\
\mathbf{E}[Y_{n+1}|T_n] &= Y_n + f'(2\alpha - 1)\left(\frac{2S_n}{n+1} - 1\right),
\end{aligned}$$

on déduit

$$\begin{aligned}
s_{n+1} &= \left(1 + \frac{2\alpha - 1}{n+1}\right) s_n + 1 - \alpha, \\
y_{n+1} &= y_n + f'(2\alpha - 1)\left(\frac{2s_n}{n+1} - 1\right).
\end{aligned}$$

Puis en tirant le fil on obtient successivement

$$\begin{aligned}
\mathbf{E}\left[\Delta X_i|T_{i-1}\right] &= \mathbf{E}\left[\Delta Y_i|T_{i-1}\right] - \Delta y_i \\
&= f'(2\alpha - 1)\left(\frac{2S_{i-1}}{i} - 1\right) + f'(2\alpha - 1)\left(\frac{2s_{i-1}}{i} - 1\right) \\
&= f'(2\alpha - 1)\frac{2(S_{i-1} - s_{i-1})}{i} \\
&= 2f'(2\alpha - 1)\frac{Z_{i-1}}{i}, \\
\Delta M_i^X &= \Delta X_i - \mathbf{E}\left[\Delta X_i|T_{i-1}\right] \\
&= \Delta X_i - 2f'(2\alpha - 1)\frac{Z_{i-1}}{i}, \\
\mathbf{E}\left[\Delta Z_i|T_{i-1}\right] &= \mathbf{E}\left[\Delta S_i|T_{i-1}\right] - \Delta s_i \\
&= \frac{2\alpha - 1}{i} S_{i-1} + 1 - \alpha - \left(\frac{2\alpha - 1}{i} s_{i-1} + 1 - \alpha\right) \\
&= \frac{2\alpha - 1}{i}\left(S_{i-1} - s_{i-1}\right) \\
&= \frac{2\alpha - 1}{i} Z_{i-1},
\end{aligned}$$

$$\begin{aligned}\Delta M_i^Z &= \Delta Z_i - \mathbf{E}\left[\Delta Z_i | T_{i-1}\right] \\ &= \Delta Z_i - \frac{2\alpha-1}{i} Z_{i-1} \\ &= Z_i - \left(1 + \frac{2\alpha-1}{i}\right) Z_{i-1}.\end{aligned}$$

Soient maintenant deux réels fixés α_1 et α_2. Pour tout n nous voulons des réels b_{in} et c_{in}, pour $i = 1, \cdots, n$ tels que

$$\sum_{i=1}^{n} b_{in} \Delta M_i^X + \sum_{i=1}^{n} c_{in} \Delta M_i^Z = \alpha_1 X_n + \alpha_2 Z_n,$$

soit

$$\begin{aligned}\sum_{i=1}^{n} b_{in} \left(\Delta X_i - 2f'(2\alpha-1)\frac{Z_{i-1}}{i}\right) + \sum_{i=1}^{n} c_{in} \left(Z_i - \left(1 + \frac{2\alpha-1}{i}\right) Z_{i-1}\right) &= \alpha_1 X_n + \alpha_2 Z_n, \\ \sum_{i=1}^{n} b_{in} \left(X_i - X_{i-1}\right) + \sum_{i=1}^{n} c_{in} Z_i - \left(c_{in}\left(1 + \frac{2\alpha-1}{i}\right) + b_{in} 2f'(2\alpha-1)\frac{1}{i}\right) Z_{i-1} &= \alpha_1 X_n + \alpha_2 Z_n.\end{aligned}$$

Après une transformation d'Abel cetté égalité devient :

$$\begin{aligned}& b_{nn} X_n + \sum_{i=1}^{n-1} \left(b_{in} - b_{i+1,n}\right) X_i + c_{nn} Z_n \\ & + \sum_{i=1}^{n-1} \left(c_{in} - c_{i+1,n}\left(1 + \frac{2\alpha-1}{i+1}\right) - b_{i+1,n} 2f'(2\alpha-1)\frac{1}{i+1}\right) Z_i = \alpha_1 X_n + \alpha_2 Z_n.\end{aligned}$$

Par identification, on trouve $b_{in} = \alpha_1$ pour tout i, et $c_{nn} = \alpha_2$,

$$c_{in} = \left(1 + \frac{2\alpha-1}{i+1}\right) c_{i+1,n} + 2f'\left(2\alpha-1\right)\frac{\alpha_1}{i}.$$

Posons

$$z_{in} = b_{in} \Delta M_i^X + c_{in} \Delta M_i^Z.$$

Il vient

$$z_{in}^2 = b_{in}^2 \left(\Delta M_i^X\right)^2 + c_{in}^2 \left(\Delta M_i^Z\right)^2 + b_{in} c_{in} \Delta M_i^X \Delta M_i^Z$$

Nous voulons trouver la limite en probabilité de $\mathbf{E}\left[z_{in}^2 | T_{i-1}\right]$. Pour cela, nous allons d'abord commencer par étudier chacun des trois termes $\left(\Delta M_i^X\right)^2$, $\left(\Delta M_i^Z\right)^2$ et $\Delta M_i^X \Delta M_i^Z$.

Premier terme $\left(\Delta M_i^X\right)^2$

Nous avons

$$\mathbf{E}\left[\left(\Delta M_i^X\right)^2 | T_{i-1}\right] = \mathbf{E}\left[\left(\Delta X_i\right)^2 | T_{i-1}\right] - \mathbf{E}\left[\Delta X_i | T_{i-1}\right]^2$$

Étudions $\mathbf{E}\left[\left(\Delta X_i\right)^2 | T_{i-1}\right]$:

$$\begin{aligned}\mathbf{E}\left[\left(\Delta X_i\right)^2 | T_{i-1}\right] &= \mathbf{E}\left[\left(\Delta Y_i - \Delta y_i\right)^2 | T_{i-1}\right] \\ &= \mathbf{E}\left[\left(\Delta Y_i\right)^2 | T_{i-1}\right] - 2\Delta y_i \mathbf{E}\left[\Delta Y_i | T_{i-1}\right] + \left(\Delta y_i\right)^2\end{aligned}$$

Or, d'une part, $y_n = f'\frac{\Gamma(n+2\alpha)}{\Gamma(2\alpha)n!} = f'\frac{n+2\alpha-1}{n}\frac{\Gamma(n+2\alpha)}{\Gamma(2\alpha)n!} = \frac{n+2\alpha-1}{n}y_{n-1}$, donc

$$\begin{aligned} y_n - y_{n-1} &= \left(\frac{n+2\alpha-1}{n} - 1\right) y_{n-1} \\ &= \frac{2\alpha-1}{n} y_n = f'\binom{i+2\alpha}{i} \\ &\sim \frac{2\alpha-1}{\Gamma(2\alpha)} n^{2\alpha-2} \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

si bien que

$$(y_n - y_{n-1})\mathbf{E}\left[(Y_n - Y_{n-1})\middle|T_{n-1}\right] \underset{n\to\infty}{\longrightarrow} 0,$$

ce qui reste bien sûr vrai pour $(y_i - y_{i-1})^2$: $(y_i - y_{i-1})^2 \underset{i\to\infty}{\longrightarrow} 0$.

D'autre part,

$$\begin{aligned} \mathbf{E}\left[\Delta Y_i | T_{i-1}\right] &= \mathbf{E}\left[Y_0\left(2\Delta S_i - 1\right)\middle|T_{i-1}\right] \\ &= \mathbf{E}\left[Y_0\right]\left(2\mathbf{E}\left[\Delta S_i\middle|T_{i-1}\right] - 1\right) \\ &\leqslant \left|\mathbf{E}\left[Y_0\right]\right|, \end{aligned}$$

donc $\Delta y_i \mathbf{E}\left[\Delta Y_i | T_{i-1}\right] \underset{i\to\infty}{\xrightarrow{P}} 0$. (Nous calculerons la limite en probabilité de $\mathbf{E}\left[\Delta Y_i | T_{i-1}\right]$ plus tard).

Par ailleurs, $\Delta Y_i = Y_0\left(2\left(\mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_<\right) - 1\right)$ donc, comme $\mathbb{1}^2_{\text{évén}^t} = \mathbb{1}_{\text{évén}^t}$ et que $\mathbb{1}_+ \cdot \mathbb{1}_- = \mathbb{1}_< \cdot \mathbb{1}_> = 0$, on a

$$(\Delta Y_i)^2 = Y_0^2\left(4\left(\mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_<\right) - 4\left(\mathbb{1}_+\mathbb{1}_> + \mathbb{1}_-\mathbb{1}_<\right) + 1\right) = Y_0^2$$

et $\mathbf{E}\left[(\Delta Y_i)^2 | T_{i-1}\right] = \mathbf{E}\left[Y_0^2\right] = f'' + f'$.

Étudions maintenant le terme $\mathbf{E}\left[\Delta X_i | T_{i-1}\right]^2$. Il est donné par

$$\mathbf{E}\left[\Delta X_i | T_{i-1}\right]^2 = f'^2 4(2\alpha-1)^2 \frac{Z_i^2}{i^2}.$$

La variance de Z_i est linéaire, donc par l'inégalité de Tchebychev, Z_i/i tend vers 0 en probabilité.

Finalement,

$$\mathbf{E}\left[\left(\Delta M_i^X\right)^2 | T_{i-1}\right] \underset{i\to\infty}{\xrightarrow{P}} f'' + f'.$$

Deuxième terme $\mathbf{E}\left[\left(\Delta M_i^Z\right)^2\right]$

Nous avons

$$\mathbf{E}\left[\left(\Delta M_i^Z\right)^2 | T_{i-1}\right] = \mathbf{E}\left[(\Delta Z_i)^2 | T_{i-1}\right] - \mathbf{E}\left[\Delta Z_i | T_{i-1}\right]^2.$$

Étudions $\mathbf{E}\left[(\Delta S_i)^2 | T_{i-1}\right]$, qui s'écrit

$$\begin{aligned} \mathbf{E}\left[(\Delta Z_i)^2 | T_{i-1}\right] &= \mathbf{E}\left[(\Delta S_i - \Delta s_i)^2 | T_{i-1}\right] \\ &= \mathbf{E}\left[(\Delta S_i)^2 | T_{i-1}\right] - 2\Delta s_i \mathbf{E}\left[\Delta S_i | T_{i-1}\right] + (\Delta s_i)^2. \end{aligned}$$

Nous avons d'une part $u_i - u_{i-1} = y_i - y_{i-1} + \frac{1}{2}$, donc $s_i - s_{i-1} \xrightarrow[i\to\infty]{P} \frac{1}{2}$, et d'autre part

$$\begin{aligned}
\mathbf{E}\left[\Delta S_i | T_{i-1}\right] &= \frac{2\alpha-1}{i} S_{i-1} + 1 - \alpha \\
&= \frac{2\alpha-1}{i}\left(S_{i-1} - a_{i-1}\right) + \frac{2\alpha-1}{i} a_{i-1} + 1 - \alpha \\
&= \frac{2\alpha-1}{i} Z_{i-1} + \frac{2\alpha-1}{i}\left(\frac{i}{2} + y_{i-1}\right) + 1 - \alpha \\
&= (2\alpha-1)\frac{Z_{i-1}}{i} + (2\alpha-1)\left(\frac{1}{2} + \frac{y_{i-1}}{i}\right) + 1 - \alpha,
\end{aligned}$$

d'où $\mathbf{E}\left[\Delta S_i | T_{i-1}\right] \xrightarrow[i\to\infty]{P} (2\alpha-1)\frac{1}{2} + 1 - \alpha = \frac{1}{2}$.

On a $\Delta S_i \in \{0,1\}$ donc $\mathbf{E}\left[\left(\Delta S_i\right)^2 | T_{i-1}\right] = \mathbf{E}\left[\Delta S_i | T_{i-1}\right]$.

Enfin, $s_i - s_{i-1} = y_i - y_{i+1} + \frac{1}{2}$ donc $s_i - s_{i-1} \xrightarrow[i\to\infty]{P} \frac{1}{2}$.

Finalement

$$\mathbf{E}\left[\left(\Delta Z_i\right)^2 | T_{i-1}\right] \xrightarrow[i\to\infty]{P} \frac{1}{2} - 2\frac{1}{2}\cdot\frac{1}{2} + \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Étudions maintenant $\mathbf{E}\left[\Delta Z_i | T_{i-1}\right]^2$. Nous avons vu

$$\mathbf{E}\left[\Delta Z_i | T_{i-1}\right] = (2\alpha-1)\frac{Z_i}{i},$$

donc

$$\mathbf{E}\left[\Delta Z_i | T_{i-1}\right] \xrightarrow[i\to\infty]{P} 0.$$

Finalement

$$\mathbf{E}\left[\left(\Delta M_i^Z\right)^2 | T_{i-1}\right] \xrightarrow[i\to\infty]{P} \frac{1}{4}.$$

Troisième terme $\mathbf{E}\left[\Delta M_i^X \cdot \Delta M_i^Z | T_{i-1}\right]$

Le terme $\mathbf{E}\left[\Delta M_i^X \cdot \Delta M_i^Z | T_{i-1}\right]$ peut se réécrire en

$$\mathbf{E}\left[\Delta M_i^X \cdot \Delta M_i^Z | T_{i-1}\right] = \mathbf{E}\left[\Delta X_i \cdot \Delta Z_i | T_{i-1}\right] - \mathbf{E}\left[\Delta X_i | T_{i-1}\right] \cdot \mathbf{E}\left[\cdot \Delta Z_i | T_{i-1}\right]$$

avec

$$\begin{aligned}
\mathbf{E}\left[\Delta X_i | T_{i-1}\right] &= 2f'(2\alpha-1)\frac{Z_{i-1}}{i}, \\
\mathbf{E}\left[\Delta Z_i | T_{i-1}\right] &= \frac{2\alpha-1}{i} Z_{i-1},
\end{aligned}$$

donc $\mathbf{E}\left[\Delta X_i | T_{i-1}\right] \cdot \mathbf{E}\left[\Delta Z_i | T_{i-1}\right] \xrightarrow[i\to\infty]{P} 0$.

De plus,

$$\begin{aligned}
\mathbf{E}\left[\Delta X_i \cdot \Delta Z_i | T_{i-1}\right] &= \mathbf{E}\left[\left(\Delta Y_i - \Delta y_i\right) \cdot \left(\Delta S_i - \Delta s_i\right) | T_{i-1}\right] \\
&= \mathbf{E}\left[\Delta Y_i \cdot \Delta S_i\right] - \Delta y_i \mathbf{E}\left[\Delta S_i | T_{i-1}\right] - \Delta s_i \mathbf{E}\left[\Delta Y_i | T_{i-1}\right] + \Delta y_i \Delta s_i
\end{aligned}$$

De $\lim P_{i\to\infty} \mathbf{E}\left[\Delta S_i | T_{i-1}\right] = \frac{1}{2}$, on déduit que $\mathbf{E}\left[\Delta Y_i | T_{i-1}\right] = \mathbf{E}\left[Y_0\right]\left(2\mathbf{E}\left[\Delta S_i | T_{i-1}\right] - 1\right)$ converge en probabilité vers 0.

En outre,

$$\begin{aligned}\mathbf{E}\left[\Delta Y_i \cdot \Delta S_i\right] &= \mathbf{E}\left[Y_0\left(2\Delta S_i - 1\right) \cdot \Delta S_i\right] \\ &= \mathbf{E}\left[Y_0 \Delta S_i\right] \\ &= \mathbf{E}\left[Y_0\right]\mathbf{E}\left[\Delta S_i\right] \\ &\xrightarrow[i\to\infty]{P} \frac{1}{2}f'.\end{aligned}$$

Les autres termes convergent en probabilité vers 0, donc finalement

$$\mathbf{E}\left[\Delta M_i^X \cdot \Delta M_i^Z | T_{i-1}\right] \xrightarrow[i\to\infty]{P} \frac{1}{2}f'.$$

Étude finale de $\mathbf{E}\left[z_{in}^2 | T_{i-1}\right]$

Des paragraphes précédents on déduit

$$\mathbf{E}\left[z_{in}^2 | T_{i-1}\right] \underset{i\to\infty}{\overset{P}{\sim}} b_{in}^2\left(f'' + f'\right) + c_{in}^2\frac{1}{4} + b_{in}c_{in}\frac{1}{2}f'.$$

Si on pose $\mathbf{c} = (\alpha_1, \alpha_2)$, alors pour $n \geqslant 1$

$$\sum_{i=1}^{n}\mathbf{E}\left[z_{in}^2\right] = \mathbf{c}^{\mathrm{T}}\Sigma_n\mathbf{c},$$

où Σ_n est la matrice de covariance de (X_n, Z_n). Si on définit

$$\Sigma_0 = \begin{bmatrix} \lambda_0 & \frac{1-C_\alpha}{2} \\ \frac{1-C_\alpha}{2} & \frac{\lambda_0}{4} \end{bmatrix},$$

qui par construction est définie positive, et

$$L_n = n\left(\mathbf{c}^{\mathrm{T}}\Sigma_0\mathbf{c}\right),$$

alors on a

$$\frac{1}{\mathbf{c}^{\mathrm{T}}\Sigma_n\mathbf{c}}\sum_{i=1}^{n}\mathbf{E}\left[z_{in}^2 | T_{i-1}\right] \xrightarrow[i\to\infty]{P} 1,$$

soit

$$\frac{1}{L_n}\sum_{i=1}^{n}\mathbf{E}\left[z_{in}^2 | T_{i-1}\right] \xrightarrow[i\to\infty]{P} 1.$$

De l'expression $c_{in} = \left(1 + \frac{2\alpha-1}{i+1}\right)c_{i+1,n} + 2f'\left(2\alpha - 1\right)\frac{\alpha_1}{i}$, on déduit

$$c_{in} \leqslant c_{i+1,n} + \frac{K}{i},$$

avec $K = 2f'\left(2\alpha - 1\right)\alpha_1$, donc $C_{in} = O(\log n)$. De plus, pour tout i, $b_{in} = \alpha_1$, donc $b_{in} = O(1)$. Il s'ensuit

$$\max_{i\leqslant n}\frac{\mathbf{E}\left[z_{in}^2 | T_{i-1}\right]}{L_n} \xrightarrow[i\to\infty]{P} 0,$$

ce qui prouve la condition de Lindeberg. □

Remarquons que si $\alpha > 1/2$, alors $\frac{c_{in}}{c_{i+1,n}} > 1 + \frac{2\alpha-1}{i+1}$, et

$$c_{1n} > \prod_{i=1}^{n} \left(1 + \frac{2\alpha - 1}{i+1}\right) = \Omega(n),$$

et la condition de Lindeberg n'est plus satisfaite.

Remarque 3.1 La variable $\widetilde{S}_n := S_n - \mathbf{E}[S_n]$ (où, rappelons-le, S_n dénombre les nœuds externes remontant positivement à la racine,) vérifie la relation de récurrence

$$\mathbf{E}[\widetilde{S}_{n+1}|T_n] = \left(1 + \frac{2\alpha - 1}{n+1}\right) \widetilde{S}_n,$$

qui rappelle beaucoup la relation de récurrence concernant le polynôme de niveau $W_n(z)$ [30] d'un arbre de taille n, dont les coefficients $[z^k]W_n(z)$ comptent le nombre de noeuds internes au niveau k :

$$\mathbf{E}[W_{n+1}(z)|T_n] = \left(1 + \frac{2z - 1}{n+1}\right) W_n(z).$$

Or le polynôme de niveau n'admet pas les même lois limites que $\widetilde{S}_n$. On voit donc que ces récurrences sont très sensibles aux premiers termes, déterminants dans la valeur asymptotique des premiers moments et la nature de la loi limite.

3.4 Conclusion

Lorsqu'on peut mettre le paramètre étudié sous la forme d'un modèle d'urnes, on peut faire apparaître une martingale et parfois montrer une limite gaussienne en utilisant le théorème central limite des martingales.

C'est ce qu'on a pu faire sur les arbres de Catalan en utilisant l'algorithme de Rémy, et sur les arbres binaires de recherche. Pour étudier la valuation sur les arbres binaires de recherche, il a fallu introduire un autre paramètre, à savoir le nombre de feuilles positives, et étudier la distribution jointe entre ce dernier paramètre et la valuation de l'arbre.

Reste qu'introduire des martingales sur des arbres de Catalan semble moins naturel que sur des arbres suivant la distribution des arbres binaires de recherche. En effet, étudier les nombres de Narayana revient à étudier la valuation des arbres de Catalan avec opérateurs $-$, d'après la bijection qu'on a montrée dans le chapitre 1. Cependant, nous n'avons pas pu introduire de martingales sur les arbres de Catalan pour le cas générique $\alpha \in]0, 1[$.

Notons qu'avec l'algorithme de Rémy, c'est la première fois qu'on fait apparaître des martingales sur des familles simples d'arbres. Une question serait de savoir si l'on peut en faire apparaître d'autres à l'aide de règles de succession plus générales [50], avec l'opérateur ECO. Suivant les règles de succession considérées, cela peut nécessiter de gérer un autre paramètre (par exemple le nombre de pas dans la dernière descente) : si ce paramètre est sympathique par rapport au paramètre auquel on s'intéresse initialement, cela pourrait permettre d'en déterminer la distribution limite.

Chapitre 4

Simplification et réduction d'arbres p-aires

Les sections 4.2 et 4.3 de ce chapitre reprennent l'article [171] en généralisant à des symboles, choisis dans un ensemble fini, pouvant être choisis avec des probabilités fixées quelconques, au lieu d'être choisis de manière uniforme. Afin que ce chapitre se suffise à lui-même, les preuves ici sont complètes, et ne s'appuient pas sur des résultats issus des articles [26, 27] ou de la thèse [64]. Elles en reprennent cependant les idées, en particulier, de [64]. Néanmoins les démonstrations de [64] ne se généralisent pas naturellement à des probabilités quelconques aux feuilles, et beaucoup de tatônnements ont été nécessaires pour trouver des inégalités du type de celles des lemmes 4.3, 4.9, 4.10 et 4.11.

La section 4.4 est nouvelle : elle traite du même problème mais avec des types d'opérateurs différents.

4.1 Introduction

Plusieurs analyses, surtout limitées à la moyenne, ont été effectuées concernant les algorithmes en calcul symbolique : différentiation symbolique [90] et des règles plus générales de réécriture [33], unification [3].

Ce chapitre porte sur l'analyse de certaines règles de réécriture, comme celles qui sont utilisées en calcul formel. Des exemples de telles règles de réécriture communément rencontrées en mathématiques sont par exemple $x^2 = x$ dans les anneaux booléens, $p \cdot x = 0$ et $x^p = x$ dans des corps de caractéristique p, $x \cdot 0 = 0$, $x \cdot 1 = 1$, $x - x = 0$ pour des réels, etc.

Nous considèrerons trois types de règles de réécriture simplifiantes.

Concernant le premier type, nous considérons des expressions engendrées par une algèbre finie partant d'un nombre fini $q \geq 2$ de constantes e_i apparaissant avec probabilité π_i, et une loi p-aire idempotente $\odot$ ($p \geq 2$) : étant donné une expression f, l'expression $\odot(f, \cdots, f)$ (p fois) est équivalente à f. L'opération de réécriture d'une expression vers son expression équivalente la plus courte sera appelée *simplification*. Pour le cas binaire, nous utiliserons la notation infixe $f \odot f$ au lieu de $\odot(f, f)$.

Concernant le second type, nous considérons le même genre d'expressions, mais cette fois la loi $\odot$ sera nilpotente, de telle sorte qu'il existe une étiquette spéciale e telle que toute expression $\odot(f, \cdots, f)$ (p fois) est équivalente à e. L'opération de réécriture d'une expression vers son expression équivalente la plus courte sera appelée *réduction*.

Concernant le troisième type, nous considérons un jeu d'opérateurs booléens de type absorption, comme X OU VRAI $\rightarrow VRAI$.

Une question naturelle consiste à savoir le gain en espace apporté par ces différents types d'opération. L'étude de ce problème pour les deux premiers types d'opérateurs fut étudiée pour la première fois dans [27] puis développée dans [64] et [26], où l'on calcule la moyenne et la variance de la taille des arbres simplifiées et réduits et l'on montre qu'elles sont linéaires en n, en utilisant la méthode de Darboux-Pólya. Des progrès ont été faits depuis avec l'analyse de singularité et le théorème des quasi-puissances développé par Hwang [108]. Dans ce chapitre, nous montrerons concernant les deux premiers types d'opérateurs que ces nouveaux outils permettent d'établir que la distribution limite de la taille des arbres simplifiés et réduits est gaussienne. Nous montrons ensuite que le troisième type d'opérateurs mène à des limites discrètes.

4.2 Simplification d'arbres

Dans cette section, nous définissons l'opération de simplification à travers un algorithme, établissons des fonctions génératrices et montrons le caractère gaussien de la distribution limite de la taille des arbres simplifiées.

4.2.1 Algorithme et fonctions génératrices

Commençons par étudier quelques exemples de simplification comme définie dans l'introduction : $a \odot b$ ne peut pas être encore simplifié, $a \odot a$ se simplifie en a, $(a \odot b) \odot (a \odot b)$ se simplifie en $a \odot b$, $(a \odot a) \odot (a \odot b)$ se simplifie en $a \odot (a \odot b)$, $(a \odot b) \odot (b \odot a)$ ne peut pas être encore simplifié, et $(a \odot a) \odot (a \odot a)$ peut se simplifier en a à travers la suite des réécritures :

$$(a \odot a) \odot (a \odot a) \rightarrow a \odot (a \odot a) \rightarrow a \odot a \rightarrow a,$$

$$(a \odot a) \odot (a \odot a) \rightarrow (a \odot a) \odot a \rightarrow a \odot a \rightarrow a.$$

Ce système est confluent, ce qui se traduit par le fait que toute expression admet une et seulement une expression équivalente la plus courte, et que la suite de réécritures n'importe pas. Néanmoins, une bonne stratégie pour simplifier une expression consiste à utiliser la procédure *simplify* définie ci-dessous, dépendant de la procédure *equal* testant l'égalité de deux arbres. La procédure *simplify* simplifie récursivement les arbres en ordre préfixe, et transforme clairement toute arbre d'expression en son équivalent le plus petit.

```
function simplify(tree :Q) :Q;
local t : array of p Q; eq : boolean; i : integer;
    if tree.degree = 0 then simplify :=tree;
        else for i :=1 to p do
                t[i] :=simplify(tree.child[i]);
            od;
            i :=1; eq :=true;
            while i≤p-1 and eq do
                if equal(tree.child[i],tree.child[i+1]) then i :=i+1;
                else eq :=false; fi od
            if eq then simplify :=t[1];
                else simplify :=⊙(t[1],···,t[p]) fi fi.
```

Programme pour la simplification d'expressions avec une loi idempotente.

Les variables arbres consistent en trois champs : *degree* (0 pour les feuilles, p pour les nœuds internes), *info* (opérande e_i (pour $i = 1, 2, \cdots, q$), ou opérateur $\odot$), *child* tableau de p sous-arbres.

À cette procédure correspond une et une seule suite de réécritures menant à l'expression équivalente la plus courte. Par exemple, l'expression

$$\big((a \odot a) \odot (b \odot b)\big) \odot (a \odot b)$$

se simplifie en $a \odot b$ suivant la suite de réécritures

$$\big((a \odot a) \odot (b \odot b)\big) \odot (a \odot b) \to \big(a \odot (b \odot b)\big) \odot (a \odot b) \to (a \odot b) \odot (a \odot b) \to a \odot b.$$

Concernant une telle procédure, on parle d'algorithme bottom-up, parce qu'elle opère d'abord sur les feuilles, puis sur les petits sous-arbres, et ainsi de suite, en montant jusqu'à la racine.

La famille $\mathcal{Q}$ d'arbres p-aires t avec q symboles possibles aux nœuds externes, pondérés par leurs poids π_t, produit des probabilités des symboles aux feuilles, est définie par la série formelle

$$\mathcal{Q} = \sum_{i=1}^{q} \pi_i \boxed{e_i} + \odot(\mathcal{Q}, \cdots, \mathcal{Q}).$$

La fonction génératrice Q de la classe $\mathcal{Q}$ s'écrit

$$Q(z) = 1 + zQ^p(z). \tag{4.1}$$

Soit ρ la singularité dominante de Q et $\tau = Q(\rho)$. Bien entendu on peut trouver la valeur de ρ à partir de la forme close $\frac{1}{pn+1}\binom{pn+1}{n} q^{(p-1)n+1}$ des coefficients de Q, qui peuvent être calculés par exemple grâce au théorème d'inversion de Lagrange appliqué à $zQ(z^p)$, mais pour des raisons qui apparaîtront ci-dessous, nous effectuerons l'analyse à l'aide de fonctions génératrices et de l'analyse de singularité.

Tant que la fonction Q peut être définie implicitement par (4.1), elle reste analytique (quand $|z| < \rho$) ; elle n'est plus analytique quand l'équation fonctionnelle définissant Q cesses d'être régulière, c'est-à-dire quand la dérivée de l'équation fonctionnelle (4.1) par rapport à Q devient nulle : cela correspond au moment où z devient égal à ρ.

On dit souvent que la singularité dominante correspond à l'échec du théorème des fonctions implicites [81].

Soient ρ et τ les solutions du système

$$\tau = 1 + \rho\tau^p, \; 1 = p\rho\tau^{p-1},$$

dont nous déduisons

$$\rho = \frac{(p-1)^{p-1}}{p^p} \text{ et } \tau = \frac{p}{p-1}. \tag{4.2}$$

On appelle arbre *irréductible* un arbre qui ne peut être simplifié. Soit $\mathcal{I}$ la classe des arbres irréductibles pondérés définie récursivement par

$$\begin{aligned} \mathcal{I} &= \sum_{i=1}^{q} \pi_i \boxed{e_i} + \big\{\odot(t_1, \cdots, t_p) \;\big|\; \exists\, i, j \text{ such that } t_i \neq t_j\big\} \\ &= \sum_{i=1}^{q} \pi_i \boxed{e_i} + \odot(\mathcal{I}, \cdots, \mathcal{I}) - \big\{\odot(t, \cdots, t) \;\big|\; t \in I\big\}, \end{aligned}$$

de fonction génératrice $I(z) = \sum_{t\in\mathcal{I}} \pi_t z^{|t|} = \sum_{n\geq 0} I_n z^n$. Il est facile de voir que le nombre d'arbres irréductibles est petit :

Lemme 4.1 *Soit ρ_I la singularité dominante de I. Alors $\rho < \rho_I < 1$. Autrement dit le nombre d'arbres irréductibles de taille n est asymptotiquement exponentiellement petit, comparé au nombre total d'arbres de cette taille.*

Démonstration. Soit $J(z) = \sum_{t\in\mathcal{I}} \pi_t^p z^{p|t|}$. On a

$$I(z) = 1 + zI^p(z) - zJ(z),$$

et

$$I(\rho_I) = 1 + \rho_I I^p(\rho_I) - \rho_I J(\rho_I).$$

D'après le théorème des fonctions implicites, I admet une singularité en ϱ tel que $1 = p\varrho I(\varrho)^{p-1}$. Or, $I(\varrho)^{p-1} > 1$, donc $\varrho < 1$, et $\rho_I \leqslant \varrho < 1$, et en fait $\rho_I = \varrho$.

On a $\pi_t^p \leqslant \pi_t$ pour tout t, donc le rayon de convergence de J est au moins égal à $\rho_I^p > \rho_I$. Donc si on pose $\tau_I = I(\rho_I)$, on a

$$\tau_I = 1 + \rho_I \tau_I^p - \rho_I J(\rho_I),\ 1 = p\rho_I \tau_I^{p-1},$$

avec $\rho_I J(\rho_I) > 0$. Les coefficients de I sont plus petits que ceux de Q, donc $\rho_I \geqslant \rho$. En comparant avec

$$\tau = 1 + \rho\tau^p,\ 1 = p\rho\tau^{p-1},$$

on voit que ρ_I ne peut pas être égal à ρ, donc $\rho_I > \rho$.

□

Pour chaque $u \in \mathcal{Q}$, soit $s(u)$ l'arbre irréductible vers lequel u se simplifie. Définissons pour chaque arbre irréductible t la fonction génératrice $Q_t(z) = \sum_{s(u)=t} z^{|u|}$, et introduisons la fonction génératrice bivariée $X(z,y) = \sum_{t\in\mathcal{Q}} \pi_t y^{|s(t)|} z^{|t|}$. Alors le lemme suivant fournit une équation fonctionnelle :

Lemme 4.2

$$X(z,y) = 1 + yzX(z,y)^p - z\sum_{t\in\mathcal{I}} \left(y^{p|t|+1} - y^{|t|}\right) Q_t^p(z).$$

Démonstration. Nous avons les équations

$$\mathcal{Q}_e = \boxed{e} + \odot(\underbrace{\mathcal{Q}_e, \cdots, \mathcal{Q}_e}_{p}),\ \mathcal{Q}_t = \odot(\mathcal{Q}_{t_1}, \cdots, \mathcal{Q}_{t_p}) + \odot(\underbrace{\mathcal{Q}_t, \cdots, \mathcal{Q}_t}_{p}),$$

où $e \in \{e_1, \cdots, e_q\}$ et $t \in \mathcal{I}$ avec $t = \odot(t_1, \cdots, t_p)$.

Sommant les égalités $y^{|t|}Q_t(z) = yz\prod_{i=1}^{p} y^{|t_i|}Q_{t_i}(z) + zy^{|t|}Q_t(z)$, nous obtenons l'expression voulue pour $X(z,y)$. □

4.2.2 Une singularité algébrique

Cette section est destinée à montrer que la singularité $z = \rho(y)$ de la fonction génératrice bivariéee satisfait certaines propriétés, si bien que $\rho(1)/\rho(u)$ ressemble à une fonction génératrice de probabilité. Ces propriétés sont résumées dans la proposition suivante :

Proposition 4.1 *Il existe des nombres réels $z_1 > \rho$, $y_1 > 1$, et des fonctions analytiques A, B et C telles que, dans le domaine*

$$\mathcal{D} = \{(z, y) \mid |z| < z_1,\ |y| < y_1\},$$

X admet la représentation

$$X(z, y) = A(z, y) + B(z, y)\sqrt{C(z, y)}.$$

Cette proposition affirme en premier lieu que pour tout y, la singularité z est du type algébrique, et en second lieu que la fonction $y \mapsto \rho(y)$ est analytique au voisinage de 1.

Pour prouver la proposition, it suffit de montrer (comme cela sera fait dans le Lemme 4.4 en-dessous) que la fonction $(z, y) \mapsto \sum_{t\in\mathcal{I}} \left(y^{p|t|+1} - y^{|t|}\right) Q_t^p(z)$ est analytique pour (z, y) au-delà de $(\rho, 1)$. Pour ce faire, nous utiliserons le fait que les fonctions Q_t sont suffisamment petites :

Lemme 4.3 *Notons π_t la produit des probabilités des étiquettes de l'arbre t. Alors il existe $z_0 > \rho$, γ, $c > 0$, tel que $\gamma^p < \rho$ et, pour tout $t \in \mathcal{I}$, $Q_t(z_0) < c\sqrt[p]{\pi_t}\gamma^{|t|}$.*

Démonstration. La preuve s'inspire fortement des travaux effectués dans [64]. On distingue le cas binaire et le cas $p \geq 3$.
Cas $p = 2$. — On veut $Q_t(z_0) \leq \sqrt{\pi_t}c\gamma^{|t|}$ pour $c > 0$, $z_0 > 1/4 = \rho = \rho_D$, $0 < \gamma < \sqrt{\rho_I}$ (en fait on va avoir $\gamma^2 < \rho$).

On a $Q_e(\rho) = 2 - 2\sqrt{1-\pi_e} = \frac{2-2\sqrt{1-\pi_e}}{\sqrt{\pi_e}}\sqrt{\pi_e}$.

Soit E l'étiquette de plus grande probabilité.

1er cas : On suppose $\pi_E \leqslant \frac{1}{2}$. Donc pour toute étiquette e, $\pi_e \leqslant \frac{1}{2}$.

Il est clair que pour $0 < x < 1$, $1 - x < \sqrt{1-x}$, donc $1 - \sqrt{1-x} < x$. Donc $Q_e(\rho') = 2 - 2\sqrt{1-\pi_e} < 2\sqrt{\pi_e} < \sqrt{2}$. Donc il existe $\rho' \in]\rho, \rho_I[$ tel qu'on a encore $Q_e(\rho') < \sqrt{2}$.

Montrons par récurrence $Q_t(\rho') < c\sqrt{\pi_t}\gamma^{|t|}$ avec $c = \sqrt{2}$. En fait on montre plus précisément : $Q_e(\rho') < 2\sqrt{\pi_e}\sqrt{\pi_e}$, où π_e est le plus grand des π_a, et $Q_t(\rho') < \sqrt{2}\pi_t\gamma^{|t|}$ pour tout $t \neq e$. On montre au passage que le rayon de convergence de Q_t est toujours au moins égal à ρ'.

On aura $Q_t < c\sqrt{\pi_t}\gamma^{|t|}$ si $\frac{1-\sqrt{1-4z^2Q_{t_1}Q_{t_2}}}{2z} < c\sqrt{\pi_t}\gamma^{|t|}$, soit si

$$\begin{aligned}
& 1 - 2z\sqrt{\pi_t}c\gamma^{|t|} < \sqrt{1 - 4z^2Q_{t_1}Q_{t_2}} \\
\Leftarrow\ & 1 - 4z\sqrt{\pi_t}c\gamma^{|t|} + 4z^2\pi_t c^2\gamma^{2|t|} < 1 - 4z^2Q_{t_1}Q_{t_2} \\
\Leftarrow\ & -4z\sqrt{\pi_t}c\gamma^{|t|} + 4z^2\pi_t c^2\gamma^{2|t|} \leqslant -4z^2\sqrt{\pi_t}c^2\gamma^{|t|-1} \\
\Leftarrow\ & -\gamma + z\sqrt{\pi_t}c\gamma^{|t|+1} < -zc \\
\Leftarrow\ & zc(1 + \sqrt{\pi_t}\gamma^{|t|+1}) < \gamma \\
\Leftarrow\ & zc(1 + \sqrt{\pi_E}\gamma^2) < \gamma \\
\Leftarrow\ & zc(1 + \sqrt{\frac{1}{2}}\gamma^2) < \gamma.
\end{aligned}$$

Or pour $c = \sqrt{2}$, $\gamma = \frac{1}{2}$, on a

$$\frac{\gamma}{c\left(1+\sqrt{\frac{1}{2}}\gamma^2\right)} = \frac{\frac{1}{2}}{\sqrt{2}\left(1+\sqrt{\frac{1}{2}}\frac{1}{4}\right)} > \frac{1}{4}.$$

Donc on peut trouver $\gamma < \frac{1}{2}$ et $z > \frac{1}{4}$ tels que l'inégalité $zc(1+\sqrt{\frac{1}{2}}\gamma^2) < \gamma$ soit encore vérifiée.

2nd cas : On suppose $\pi_E > \frac{1}{2}$. Les autres π_e ($e \neq E$) sont plus petits que $1-\pi_E$.

Soit $\varepsilon > 0$ un réel arbitrairement petit. On a $Q_E(\rho') < \frac{2-2\sqrt{1-\pi_E}}{\sqrt{\pi_E}}\sqrt{\pi_E}$. La fonction $x \mapsto \frac{1-\sqrt{1-x}}{\sqrt{x}}$ est strictement croissante, et pour $e \neq E$ $Q_e(\rho) = \frac{2-2\sqrt{1-\pi_e}}{\sqrt{\pi_e}}\sqrt{\pi_e} \leqslant \frac{2-2\sqrt{\pi_M}}{1-\sqrt{\pi_M}}\sqrt{\pi_e}$.

Soit $c_E = \frac{2-2\sqrt{1-\pi_E}}{\sqrt{\pi_E}} + \varepsilon$. Montrons par récurrence $Q_t(\rho') < c\sqrt{\pi_t}\gamma^{|t|}$ avec $c > \frac{2-2\sqrt{\pi_E}}{\sqrt{1-\pi_E}}$ si $t \neq E$, et $\gamma \neq \frac{1}{2}$.Pour ρ' légèrement plus grand que ρ, et strictement plus petit que ρ_I. on a $Q_E(\rho') < c_E\sqrt{\pi_E}$.

On a $Q_t = zQ_{t_1}Q_{t_2} + zQ_t^2 = \frac{1-\sqrt{1-4z^2Q_{t_1}Q_{t_2}}}{2z}$, où l'un au moins des sous-arbres est différent de E. Autrement dit, par hypothèse de récurrence, $Q_{t_1} < c_1\sqrt{\pi_{t_1}}\gamma^{|t_1|}$ et $Q_{t_2} < c_2\sqrt{\pi_{t_2}}\gamma^{|t_2|}$, où l'un au moins des c_i est plus petit que c_E, et l'autre plus petit que c.

On aura $Q_t < c\sqrt{\pi_t}\gamma^{|t|}$ si $\frac{1-\sqrt{1-4z^2Q_{t_1}Q_{t_2}}}{2z} < c\sqrt{\pi_t}\gamma^{|t|}$, soit si

$$\begin{aligned}
& 1 - 2z\sqrt{\pi_t}c\gamma^{|t|} < \sqrt{1-4z^2Q_{t_1}Q_{t_2}} \\
\Leftarrow\ & 1 - 4z\sqrt{\pi_t}c\gamma^{|t|} + 4z^2\pi_t c^2\gamma^{2|t|} < 1 - 4z^2Q_{t_1}Q_{t_2} \\
\Leftarrow\ & -4z\sqrt{\pi_t}c\gamma^{|t|} + 4z^2\pi_t c^2\gamma^{2|t|} \leqslant -4z^2\sqrt{\pi_t}cc_E\gamma^{|t|-1} \\
\Leftarrow\ & zc_E \leqslant \gamma - z\sqrt{\pi_t}c\gamma^{|t|+1} \\
\Leftarrow\ & zc_E \leqslant \gamma - z\sqrt{\pi_t}c\gamma^2.
\end{aligned}$$

Pour $\gamma = \frac{1}{2}$ cela est impliqué par

$$\begin{aligned}
& z < \frac{2}{4c_E + \sqrt{\pi_t}c} \\
\Leftarrow\ & z < \frac{2}{4c_E + \sqrt{\pi_E}c}.
\end{aligned}$$

On veut pouvoir choisir $\frac{1}{4} < z$, donc on veut

$$\begin{aligned}
& \frac{1}{4} < \frac{2}{4c_E + \sqrt{\pi_E}c} \\
\Leftarrow\ & 4c_E + \sqrt{\pi_E}c < 8 \\
\Leftarrow\ & 4c_E + \sqrt{\pi_E}c < 8 \\
\Leftarrow\ & 8 - 4\frac{2-2\sqrt{1-\pi_E}}{\sqrt{\pi_E}} - \sqrt{\pi_E}\frac{2-2\sqrt{\pi_E}}{\sqrt{1-\pi_E}} > 0.
\end{aligned}$$

Or il est facile de vérifier que la fonction $x \mapsto 8 - 4\frac{2-2\sqrt{1-x}}{\sqrt{x}} - \sqrt{x}\frac{2-2\sqrt{x}}{\sqrt{1-x}}$ est strictement positive sur $]1/2, 1]$. Donc on a $z < \frac{2}{4c_E+\sqrt{\pi_t}c}$ pour $z = 1/4$, $c_E = \frac{2-2\sqrt{1-\pi_E}}{\sqrt{\pi_E}}$ et $c = \frac{2-2\sqrt{\pi_E}}{\sqrt{1-\pi_E}}$, donc il existe $\varepsilon > 0$ tel que cette égalité reste vérifiée si $z \leqslant 1/4 + \varepsilon$, $c_E \leqslant \frac{2-2\sqrt{1-\pi_E}}{\sqrt{\pi_E}} + \varepsilon$, et $c \leqslant \frac{2-2\sqrt{\pi_E}}{\sqrt{1-\pi_E}} + \varepsilon$.

Pour tout $0 < x < 1$, $1-\sqrt{1-x} < x$ (car $1-x < \sqrt{1-x}$), donc $c\gamma < 16z^2\sqrt{\pi_e\pi_t}/(2z\sqrt{\pi_t}) = 8z\sqrt{\pi_e}$, et on aura $\gamma < \frac{1}{2}$ si

$$\begin{aligned}
& 1-\sqrt{1-16z^2\sqrt{\pi_e\pi_t}} < z\sqrt{\frac{\pi_t}{2}} \\
\Leftarrow\ & 1-\sqrt{\frac{\pi_t}{2}}z < \sqrt{1-16z^2\sqrt{\pi_e\pi_t}} \\
\Leftarrow\ & 1-\sqrt{2\pi_t}z+\frac{\pi_t}{2}z^2 < 1-16z^2\sqrt{\pi_e\pi_t} \\
\Leftarrow\ & z^2\left(\frac{\pi_t}{2}+16z^2\sqrt{\pi_e\pi_t}\right) < \sqrt{2\pi_t}z \\
\Leftarrow\ & z < \frac{\sqrt{2\pi_t}}{\frac{\pi_t}{2}+16\sqrt{\pi_e\pi_t}} = \frac{\sqrt{2}}{\frac{\sqrt{\pi_t}}{2}+16\sqrt{\pi_t}} \\
\Leftarrow\ & z < \frac{\sqrt{2}}{1+16}.
\end{aligned}$$

Cas $p \geq 3$. — On veut

$$Q_t(z_0) \leqslant c\sqrt[p]{\pi_t}\gamma^{|t|},\ z_0 > \rho,\ 0 < \gamma < \sqrt[p]{\rho_I}.$$

On a la relation de récurrence

$$Q_t(z_0) - z_0 Q_t^p(z_0) = z_0 \prod_{i=1}^{p} Q_{t_i}(z_0) < z_0 c\sqrt[p]{\pi_t}\gamma^{|t|}.$$

On introduit la fonction $g(x) = x - z_0 x^p$. $Q_t(z_0)$ est la plus petite racine positive de l'équation $g\left(Q_t(z_0)\right) = z_0 \prod_{i=1}^{p} Q_{t_i}(z_0)$: utiliser par exemple le théorème des fonctions implicites. (Ou montrer que l'autre solution réelle correspond à une série entière non définie en $z = 0$).

Cela implique en particulier que $Q_t(z_0)$ est plus petit que x_0, valeur en laquelle g atteint son maximum. Or $g_{|]-\infty,x_0]}$ est croissante, si bien que pour montrer $Q_t(z_0) \leqslant c\sqrt[p]{\pi_t}\gamma^{|t|}$, il suffit de montrer $g\left(Q_t(z_0)\right) < g\left(c\sqrt[p]{\pi_t}\gamma^{|t|}\right)$. Or par hypothèse de récurrence, $g\left(Q_t(z_0)\right) < z_0 c\sqrt[p]{\pi_t}\gamma^{|t|}$, donc il suffit de montrer

$$\begin{aligned}
c\sqrt[p]{\pi_t}\gamma^{|t|} &\leqslant g\left(c\sqrt[p]{\pi_t}\gamma^{|t|}\right) \\
&\leqslant c\sqrt[p]{\pi_t}\gamma^{|t|} - z_0\left(c\sqrt[p]{\pi_t}\gamma^{|t|}\right)^p,
\end{aligned}$$

soit pour tout $|t| \geqslant 1$,

$$z_0 c^{p-1}\left(\frac{1}{\gamma}+\gamma^{(p-1)|t|}\right) < 1.$$

Or $\gamma < 1$, donc l'égalité sera vraie en général si elle est déjà vraie pour $|t| = 1$, soit

$$z_0 c^{p-1}\left(\frac{1}{\gamma}+\gamma^{p-1}\right) < 1.$$

Notons que pour tout i,

$$\rho_i = \rho_{Q_i} = \frac{(p-1)^{p-1}}{p^p \pi_i^{p-1}},\ Q_i(\rho_i) = \frac{p\pi_i}{p-1}.$$

Soit i_0 tel que $\pi_{i_0} = \max_i \pi_i$. Posons $c = \frac{p}{p-1}\pi_{i_0}^{1-\frac{1}{p}}$. Alors pour tout i, $Q_i(\rho) < c\pi_i^{\frac{1}{p}}$. Rappelons $\rho = (p-1)^{p-1}p^{-p}$. Considérons la fonction $H(z_0,\gamma) = z_0\left(\frac{p}{p-1}\right)^{p-1}\frac{1}{\gamma}(1+\gamma^p)$. On a :

$$\begin{aligned} H(\rho,\rho^{\frac{1}{p}}) &= \rho\left(\frac{p}{p-1}\right)^{p-1}\frac{1}{\rho^{\frac{1}{p}}}(1+\rho^p) \\ &= \left(\frac{p}{p-1}\right)^{p-1}\rho^{1-\frac{1}{p}}(1+\rho^p) \\ &= \left(\frac{p}{p-1}\right)^{p-1}\left(\frac{(p-1)^{p-1}}{p^p}\right)^{1-\frac{1}{p}}\left(1+\left(\frac{(p-1)^{p-1}}{p^p}\right)^p\right) \end{aligned}$$

Montrons que $H(\rho,\rho^{\frac{1}{p}}) < 1$ pour $p \geqslant 3$. Nous avons

$$\begin{aligned} H(\rho,\rho^{\frac{1}{p}}) &= p^{p-1}p^{-p\frac{p-1}{p}}(p-1)^{-(p-1)+(p-1)-\frac{1}{p}(p-1)}\left(1+\frac{(p-1)^{p^2-p}}{p^{p^2}}\right) \\ &= (p-1)^{-(1-\frac{1}{p})}\left(1+\left(1-\frac{1}{p}\right)^{p^2}(p-1)^{-p}\right) \\ &= e^{-(1-\frac{1}{p})\log(p-1)}\left(1+e^{p^2\log\left(1-\frac{1}{p}\right)}e^{-p\log(p-1)}\right) \end{aligned}$$

donc nous aurons $H(\rho,\rho^{\frac{1}{p}}) < 1$ si et seulement si

$$1+e^{p^2\log\left(1-\frac{1}{p}\right)}e^{-p\log(p-1)} < e^{\left(1-\frac{1}{p}\right)\log(p-1)}.$$

Cette inégalité est vérifiée pour $p=3$. Or $p^2\log\left(1-\frac{1}{p}\right)$ et $-p\log(p-1)$ sont strictement négatifs pour $p \geqslant 3$, si bien que le membre de gauche est inférieur à 2 pour $p \geqslant 3$. D'autre part, le membre de droite est supérieur à 2 pour $p=4$, et c'est une fonction croissante en p, puisque

$$\frac{d}{dp}\left(1-\frac{1}{p}\right)\log(p-1) = \frac{1-\frac{1}{p}}{p-1}+\frac{1}{p^2}\log(p-1) > 0.$$

L'inégalité $H(z_0,\gamma) < 1$ reste donc vraie pour (z_0,γ) restant dans un voisinage de $(\rho,\rho^{\frac{1}{p}})$. □

Lemme 4.4 *Soit* $d(z,y) = \sum_{t\in\mathcal{I}}\left(y^{p|t|+1}-y^{|t|}\right)Q_t^p(z)$ *la fonction telle que*

$$X(z,y) = 1+yzX(z,y)^p - zd(z,y).$$

Il existe $z_0 > \rho$ *et* $y_0 > 1$ *tels que* $d(z,y)$ *est holomorphe en les deux variables* z *et* y *et dans le domaine* $|z| < z_0$, $|y| < y_0$.

Démonstration. Soient z_0 et γ comme dans le Lemma 4.3. Comme on a $1 < \rho\gamma^{-p} < \rho_I\gamma^{-p}$, on peut trouver un nombre réel y_0 dans l'intervalle $]1,\rho_I^{1/p}\gamma^{-1}[$ et, pour $|z| < z_0$, $|y| < y_0$:

$$\sum_{t\in\mathcal{I}} y^{p|t|+1}Q_t^p(z) \leqslant \sum_{t\in\mathcal{I}} y_0^{p|t|+1}Q_t^p(z) \leqslant \sum_{n\geq 0} I_n y_0^{pn+1}c^p\gamma^{pn}$$

et

$$\sum_{t\in\mathcal{I}} y^{|t|}Q_t^p(z) \leqslant \sum_{t\in\mathcal{I}} y_0^{|t|}Q_t^p(z) \leqslant \sum_{n\geq 0} I_n y_0^n c^p \gamma^{pn},$$

dont on déduit que les séries numériques

$$\sum_{t\in\mathcal{I}} y^{p|t|+1}Q_t^p(z) \text{ and } \sum_{t\in\mathcal{I}} y^{|t|}Q_t^p(z)$$

convergent. Comme elles ont des coefficients positifs en $z^n y^k$, elles sont normalement convergentes, et donc holomorphes en z et y. Par conséquent, leur différence $d(z,y)$ est aussi holomorphe en z et y pour $|z| < z_0$, $|y| < y_0$. □

Nous sommes maintenant en mesure de montrer que la fonction génératrice bivariée X admet une singularité algébrique, et de prouver la Proposition 4.1.

Preuve de la Proposition 4.1. Soit $F(z,y,x) = x-1-yzx^p+zd(z,y)$, qui satisfait $F(z,y,X(z,y)) = 0$ pour tout z,y dans le domaine de convergence de X. Tant que d est analytique et que $\frac{\partial F}{\partial x}$ est non nul, le théorème des fonctions implicites affirmes que X est analytique en z et y (c'est une conséquence directe du théorème des residus qui permet d'écrire $X(z,y) = \int_C \frac{F_x(z,y,u)}{F(z,y,u)} u\,du$ pour une courbe appropriée C).

Soient z_0 et y_0 des réels comme dans le Lemme 4.4. Montrons que pour tout y tel que $|y| < y_0$, $z \mapsto X(z,y)$ admet une singularité $z = \rho(y)$ implicitement définie par une fonction analytique et qui est du type quadratique dans un voisinage de $y = 1$.

Pour tout y, le système d'équations

$$\begin{cases} F(z,y,x) &= 0 \\ \frac{\partial F}{\partial x} &= 0 \end{cases}$$

définit une singularité $\rho(y)$ dépendant de y. Ce système définit donc implicitement une fonction $z = \rho(y)$. Réduisons ce système à une seule équation. Le système peut s'écrire

$$\begin{cases} x - q - yzx^p + zd(z,y) &= 0 \quad (4.2.2.a) \\ 1 - pyzx^{p-1} &= 0. \quad (4.2.2.b) \end{cases}$$

(4.2.2.b) se réécrit en $x^{p-1} = \frac{1}{pyz}$, et le remplacement de x^p par $\frac{x}{pyz}$ dans (4.2.2.a) délivre une équation linéaire en x qui se résout en $x = \frac{1-zd}{1-\frac{1}{p}}$. En injectant cette expression dans (4.2.2.b) on trouve

$$1 - pyz\left(\frac{1-zd}{1-\frac{1}{p}}\right)^{p-1} = 0,$$

ou $C(z,y) = 0$, avec

$$C(z,y) = -\frac{1}{p}\left(1-\frac{1}{p}\right)^{p-1} + yz\big(q - zd(z,y)\big)^{p-1}.$$

D'après le Lemma 4.4 portant sur $d(z,y)$, la fonction C est analytique dans un domaine de la forme $\{\ (z,y)\ |\ |z| < z_0,\ |y| < y_0\}$, avec $z_0 > \rho$ et $y_0 > 1$. De plus, elle vérifie $C_z(\rho,1) \neq 0$. En

conséquence, la singularité $z = \rho(y)$ de $z \mapsto X(z, y)$ est implicitement définie dans un voisinage de $y = 1$ par $C(\rho(y), y) = 0$.

Comme la fonction définissant implicitement X satisfait

$$F_{x^2}(\rho, 1, X(\rho, 1)) = (p-1)p\rho X(\rho, 1)^{p-2} \neq 0,$$

$\rho(1)$ est une singularité de type quadratique, et cela reste vrai pour $\rho(y)$ si y reste au voisinage de 1. Ainsi, il existe des réels z_1, y_1, et des fonctions analytiques A et B telles que $\rho < z_1$, $1 < y_1$, et X admet une représentation

$$X(z, y) = A(z, y) + B(z, y)\sqrt{C(z, y)}.$$

□

4.2.3 Loi limite

Avant de faire appel au théorème des quasi-puissances, nous devons établir la condition de variabilité, ce qui est l'objet du lemme suivant.

Proposition 4.2 *L'espérance μ_n et la variance v_n de la taille des arbres simplifiés vérifient*

$$\mu_n = \mu n + O(1), v_n = \overline{v} n + O(1),$$

avec $\mu \in]0, 1[$ et $\overline{v} \geqslant 0$.

Démonstration. Par analyse de singularité, on a l'asymptotique

$$\mu_n = \frac{[z^n]X_y(z, 1)}{[z^n]Q(z)} = \mu n + O(1) \text{ avec } \mu = 1 - \frac{d_y(\rho, 1)}{\tau - 1},$$

et

$$v_n = \frac{[z^n]\left(X_{y^2}(z, 1) + X_y(z, 1)\right)}{[z^n]Q(z)} - \mu_n^2 = \overline{v} n + O(1).$$

L'expression de μ nous montre que μ est strictement plus petit que 1. On peut montrer que μ est non nul. Supposons en effet que l'espérance soit bornée par K.

Nous avons vu que tous les Q_t étaient analytiques au moins jusqu'à $z_0 > \rho$, donc c'est aussi le cas de

$$Q_{\leqslant \lceil K \rceil} := \sum_{t \in \mathcal{I},\ |t| \leqslant \lceil K \rceil} Q_t,$$

qui est une somme finie de Q_t. Les coefficients de $Q_{\leqslant \lceil K \rceil}$ étant positifs, il est facile de voir que, pour tout n, $[z^n]Q_{\leqslant \lceil K \rceil} < Q_{\leqslant \lceil K \rceil}(z_0)z_0^{-n}$. Comme d'autre part nous avons $[z^n]Q(z) \simeq C_Q \rho^n / \sqrt{\pi n^3}$ pour une certaine constante C_Q, cela entraîne que

$$\lim_{n \to \infty} \frac{[z^n]Q_{\leqslant \lceil K \rceil}}{[z^n]Q(z)} = 0,$$

et $\lim_{n \to \infty} \mu_n \geqslant \lceil K \rceil + 1 > K$, ce qui est absurde. □

Notons que si $\overline{v} = 0$, cela signifie que la variance est bornée. On peut donc raisonnablement penser que $\overline{v}$ est strictement positif dans le cas général, mais seuls des calculs numériques ou des simulations permettent de corroborer ce fait dans des cas particuliers.

Il est maintenant possible de trouver la distribution limite de la taille des arbres simplifiés.

Théorème 4.1 *On suppose la variance non dégénérée. Alors la taille des arbres p-aires a une distribution limite gaussienne, avec moyenne μ_n et variance v_n, telles que*

$$\mu = \lim_{n\to\infty} \frac{\mu_n}{n} = 1 - \rho(p-1)d_y(\rho,1)$$

et

$$\begin{aligned}\overline{v} &= \lim_{n\to\infty} \frac{v_n}{n} = \left[\rho(p-1) - (p-1)(p-2)\rho^2\right] d_y(\rho,1) - 3(p-1)^2\rho^2 d_y^2(\rho,1) \\ &\quad +2\rho^2(p-1)d_{yz}(\rho,1) - 2\rho^3\,(p-1)^2\, d_y(\rho,1)d_{yz}(\rho,1) - (p-1)\rho d_{y^2}(\rho,1).\end{aligned}$$

Démonstration. Selon le Lemme 4.1, dans un domaine de la forme

$$\mathcal{D} = \{(z,y) \mid |z| < z_1,\ |y| < y_1\}$$

pour un certain $z_1 > \rho$ et un certain $y > 1$, X admet la représentation

$$X(z,y) = A(z,y) + B(z,y)\sqrt{C(z,y)}$$

avec A, B et C analytique,

$$B(\rho,1) \neq 0$$

et

$$C(z,y) = -\frac{1}{p}\left(1 - \frac{1}{p}\right)^{p-1} + yz\big(1 - zd(z,y)\big)^{p-1}.$$

Comme $d(z,1) \equiv 0$, il est facile de voir que $\zeta = \rho$ est une racine simple de $\zeta \mapsto C(\zeta,1)$. X satisfait donc les conditions algebraïco-logarithmiques du théorème 9.8 de [85] p.68-69, qui établit une limite gaussienne pour toute variable aléatoire Y_n de fonction generatrice de probabilité

$$\frac{[z^n]X(z,y)}{[z^n]X(z,1)},$$

étant donné la condition de variabilité

$$\underline{\lim}_{n\to\infty} \frac{\mathrm{Var}(Y_n)}{n} > 0,$$

vérifiée par hypothèse.

On sait à partir du même théorème 9.8 que $\mu = \lim \frac{1}{n}E[Y_n]$ et $\overline{v} = \lim \frac{1}{n}\mathrm{Var}(Y_n)$ sont donnés par les expressions

$$\begin{aligned}\mu &= -\frac{\rho_y(1)}{\rho(1)}, \\ \overline{v} &= -\frac{\rho_{y^2}(1) + \rho_y(1)}{\rho(1)} + \left(\frac{\rho_y(1)}{\rho(1)}\right)^2.\end{aligned}$$

Pour calculer ces valeurs, il est seulement besoin de résoudre le système d'équations

$$C_y(\rho,1) = 0,\ C_{y^2}(\rho,1) = 0,$$

et de remarquer que $d_z(\rho,1) = d_{z^2}(\rho,1) = 0$, ce qui est plus simple que d'utiliser les dérivées partielles $X_y(z,1)$ et $X_{y^2}(z,1)$ et les égalités asymptotiques du lemme précédent.

Pour le calcul de la moyenne, on a en effet d'une part

$$\begin{aligned} C(\rho(y),y) &= 0 \\ C_y(\rho(y),y)+\rho_y(y)C(\rho(y),y) &= 0. \end{aligned}$$

et d'autre part

$$\begin{aligned} C(z,y) &= -\frac{1}{p}\left(1-\frac{1}{p}\right)^{p-1}+yz\,(1-zd(z,y))^{p-1} \\ C_y(z,y) &= z\,(1-zd(z,y))^{p-1}-(p-1)yz^2d_y(z,y)\,(1-zd(z,y))^{p-2} \\ C_z(z,y) &= y\,(1-zd(z,y))^{p-1}-(p-1)yz\,(d(z,y)+zd_y(z,y))\,(1-zd(z,y))^{p-2} \end{aligned}$$

d'où

$$\begin{aligned} C_y(\rho,1) &= \rho-\rho^2(p-1)d_y(z,1) \\ C_z(\rho,1) &= 1. \end{aligned}$$

Donc en $y=1$ (et $\rho(1)=\rho$),

$$\begin{aligned} \rho_y(1) &= -\frac{C_y(\rho,1)}{C_z(\rho,1)} \\ &= -\rho+(p-1)\rho^2d_y(z,1) \\ -\frac{\rho_y(1)}{\rho} &= 1-(p-1)\rho d_y(z,1) \end{aligned}$$

Pour le calcul de la variance, on a d'une part

$$C_{y^2}(\rho(y),y)+\rho_{y^2}C_z(\rho(y),y)+2\rho_yC_{zy}(\rho(y),y)=0,$$

et d'autre part

$$\begin{aligned} C_{y^2}(z,y) &= -2(p-1)z^2d(z,y)\,(1-zd(z,y))^{p-1}-(p-1)yz^2d_y(z,y)\,(1-zd(z,y))^{p-2} \\ &\quad +(p-1)(p-2)z^3d_y^2(z,y)\,(1-zd(z,y))^{p-3} \\ C_{zy}(z,y) &= (1-zd(z,y))^{p-1}-(p-1)yzd_y(z,y)\,(1-zd(z,y))^{p-2} \\ &\quad -(p-1)yz\,(d_y(z,y)+zd_{yz}(z,y))\,(1-zd(z,y))^{p-2} \\ &\quad -(p-1)z\,(d(z,y)+zd_z(z,y))\,(1-zd(z,y))^{p-2} \\ &\quad +(p-1)(p-2)yz\,(d(z,y)+zd_z(z,y))\,(1-zd(z,y))^{p-3} \end{aligned}$$

donc

$$\begin{aligned} \rho_{y^2} &= -\frac{C_{y^2}+2\rho_yC_{zy}}{C_z} \\ &= 2\rho^2(p-1)d_y+\rho^2(p-1)d_{y^2}-\rho^3(p-1)(p-2)d_y^2+2\rho_y\left(2\rho(p-1)d_y-1-\rho^2(p-1)d_{zy}\right) \end{aligned}$$

puis

$$\begin{aligned} \overline{v} &= -\frac{\rho_{y^2}+\rho_y}{\rho}+\left(\frac{\rho_y}{\rho}\right)^2 \\ &= \rho^2(p-1)(p-2)d_y^2-2\rho(p-1)d_y-\rho(p-1)d_{y^2} \\ &\quad -2\,(1-(p-1)\rho d_y(z,1))\,\left(1-2\rho(p-1)d_y-\rho^2(p-1)\right)d_{zy} \\ &\quad +1-(p-1)\rho d_y(z,1)+(1-(p-1)\rho d_y(z,1))^2 \end{aligned}$$

□

Des calculs numériques effectués par Fernández-Camacho dans [64] pour $p = q = 2$ et $\pi_1 = \pi_2 = 1/2$ donnent les valeurs $\mu = 0.8196...$ et $\overline{v} = 0.2166...$. Heuristiquement, Fernández-Camacho a aussi obtenu que μ et $\overline{v}$ sont respectivement croissant et décroissant lorsque p et q sont croissants.

4.3 Lois p-aires nilpotentes

Dans cette section, comme pour la simplification, nous définissons l'opération de réduction à travers un algorithme, établissons des fonctions génératrices et montrons le caractère gaussien de la distribution limite de la taille des arbres réduits.

4.3.1 Algorithme et fonctions génératrices

Nous considérons maintenant un algorithme de réduction un peu différent de la simplification. Nous travaillons maintenant avec un opérateur d'arité p, qui est nilpotent au lieu d'être idempotent : il existe un opérande particulier e tel que, si un arbre a des sous-arbres identiques, il se réduit à e.

Des exemples de réduction, comme définis dans l'introduction, sont, pour un opérateur $\odot$: $a \odot b$ ne peut être réduit, $a \odot a$ se réduit en e, $(a \odot b) \odot (a \odot b)$ se réduit en e, $(a \odot a) \odot (a \odot b)$ se réduit en $e \odot (a \odot b)$, $(a \odot b) \odot (b \odot a)$ ne peut être réduit, et $(a \odot a) \odot (a \odot a)$ peut se réduire en e à travers la suite de règles simplifiantes :

$$(a \odot a) \odot (a \odot a) \to e \odot (a \odot a) \to e \odot e \to e,$$

$$(a \odot a) \odot (a \odot a) \to (a \odot a) \odot e \to e \odot e \to e.$$

Nous définissons une procédure *reduce* de manière très similaire à la procédure *simplify* : on remplace seulement les occurrences de *simplify* par *reduce* dans l'algorithme, et on définit le résultat comme étant égal à e quand on trouve des sous-arbres identiques, au lieu d'être égal au sous-arbre commun lui-même.

À titre d'exemple, l'expression $\big((a \odot a) \odot (b \odot b)\big) \odot (a \odot b)$ se réduit en $e \odot (a \odot b)$ suivant la suite de règles

$$\big((a \odot a) \odot (b \odot b)\big) \odot (a \odot b) \to \big(e \odot (b \odot b)\big) \odot (a \odot b) \to (e \odot e) \odot (a \odot b) \to e \odot (a \odot b).$$

On peut noter que dans cet exemple la réduction de l'expression initiale mène à un arbre possédant plus de nœuds que l'arbre obtenu par simplification.

Nous considérons la classe $\mathcal{R}$ des arbres p-aires dont les nœuds externes peuvent être munis d'une des q constantes e_{i_n} dont le symbole particulier $e = e_1$, et dont les nœuds internes comportent l'opérateur nilpotent. Comme précédemment, on affecte à chaque symbole a_i une probabilité π_i.

La fonction génératrice énumérant les objets pondérés de la classe $\mathcal{R}$ par leur taille vérifie

$$R(z) = \sum_{t \in \mathcal{R}} \pi_t z^{|t|} = 1 + zR^p(z).$$

On note encore ρ la singularité dominante de R, et $\tau = R(\rho)$.

Nous définissons à nouveau la notion d'arbre *irréductible*, qui est un arbre qui ne peut être réduit. La classe des arbres irréductibles est isomorphe à la classe précédente définie dans le cas

de la simplification : dans les deux cas, un arbre est irréductible si et seulement si tout nœud interne possède au moins deux sous-arbres différents. Comme il ne peut y avoir de confusion possible, nous noterons encore cette classe $\mathcal{I}$. La fonction génératrice énumérant les objets de la classe $\mathcal{I}$ par leur taille est la même série génératrice $I(z)$ que celle qui a été introduite dans la section précédente. Soit ρ_I sa singularité dominante.

Pour tout $u \in \mathcal{R}$, soit $s(u)$ l'arbre irréductible vers lequel u se réduit. Définissons pour tout arbre irréductible t la fonction génératrice $R_t(z) = \sum_{r(u)=t} \pi_t z^{|u|}$, et introduisons la fonction génératrice bivariée $H(z,y) = \sum_{t\in\mathcal{R}} \pi_t y^{|r(t)|} z^{|t|}$. H est donné par l'équation fonctionnelle suivante :

Lemme 4.5

$$H(z,y) \;=\; 1 + yzH(z,y)^p - z\sum_{t\in\mathcal{I}} \left(y^{p|t|+1} - 1\right) R_t^p(z).$$

Démonstration. Pour tout arbre irréductible t, on note $\mathcal{R}_t$ la classe des arbres pondérés qui se réduisent en t. Nous avons :

$$\forall a \in \{e_2, \cdots, e_q\},\;\; \mathcal{R}_a \;=\; \pi_a \boxed{a},$$

$$\mathcal{R}_e \;=\; \pi_e \boxed{e} + \odot(\underbrace{\mathcal{R}_e, \cdots, \mathcal{R}_e}_{p}),$$

$$\forall t = \odot(t_1, \cdots, t_p) \in \mathcal{I},\;\; \mathcal{R}_t \;=\; \odot_i(\mathcal{R}_{t_1}, \cdots, \mathcal{R}_{t_p}),$$

ce qui, en termes de fonctions génératrices, devient :

$$\forall a \in \{e_1, \cdots, e_q\},\;\; R_a(z) = \pi_a,$$

$$R_e(z) = \pi_e + z\sum_{t\in\mathcal{I}} R_t(z)^p,$$

$$\forall t = \odot(t_1, \cdots, t_p) \in \mathcal{I},\;\; R_t(z) = z\prod_{j=1}^{p} R_{t_j}(z).$$

Sommant ces égalités, nous obtenons l'expression voulue de $H(z,y)$. □

4.3.2 Une singularité algébrique

Pour obtenir une représentation de la fonction génératrice bivariée H, nous devons montrer que la fonction $(z,y) \mapsto \sum_{t\in\mathcal{I}} \left(y^{p|t|+1} - 1\right) R_t^p(z)$ est analytique assez loin (Lemme 4.12). Pour ce faire, nous allons encore une fois montrer que R_t est analytique jusqu'à un certain $z_0 > \rho$, puis trouver des bornes supérieures pour $R_t(z_0)$.

Lemme 4.6 *Il existe $\varepsilon > 0$ tel que, pour tout $t \in \mathcal{I}$, R_t a un rayon de convergence au moins égal à ε.*

Démonstration. Pour $a \neq e$, $R_a = \pi_a$ possède un rayon de convergence infini. En déroulant la relation de récurrence $R_t(z) = z\prod_{j=1}^{p} R_{t_j}(z)$ pour $t = \odot_i(t_1, \cdots, t_p)t$ de taille au moins un, on trouve l'égalité

$$R_t(z) = \prod_{i=2}^{q} \pi_{a_i}^{|t|_{a_i}} z^{|t|} R_e(z)^{|t|_e},$$

si bien que les R_t, pour $|t| \geqslant 1$, ont même rayon de convergence que R_e. Il suffit donc de montrer que R_e converge au-delà de ρ.

R_e est implicitement défini par

$$F(z, R_e(z)) = 0$$

avec

$$F(z,y) = y - \pi_e - z \sum_{t \in \mathcal{I}} \left(\prod_{i=2}^{q} \pi_{e_i}^{|t|_{e_i}} z^{|t|} y^{|t|_e} \right)^p .$$

Les coefficients de R_e étant plus petits que ceux de R, il est clair que le rayon de convergence de R_e vaut au moins ρ. Pour montrer que R_e converge au-delà, il suffit donc de montrer

$$\frac{\partial F}{\partial y}(\rho, R_e(\rho)) > 0.$$

Or $\frac{\partial F}{\partial y}(z, R_e(z)) = 1 - pz \sum_{t \in \mathcal{I}} |t_e| R_t^p(z) R_e(z)^{-1}$. Montrons que pour tout $t \in \mathcal{I}$,

$$|t_e| R_t^p(\rho) R_e(\rho)^{-1} \leqslant R_t^{p-1}(\rho),$$

soit

$$|t_e| R_t(\rho) R_e(\rho)^{-1} \leq 1,$$

avec parfois inégalité stricte. Cette inégalité est clairement vérifiée pour $t = e$, auquel cas on a égalité, et pour $|t|_e = 0$ — on a alors une inégalité stricte. Supposons qu'on ne soit pas dans ces cas-là, ce qui implique $|t| \geqslant 1$.

Dans un arbre irréductible, tout nœud interne a au moins un fils différent de e. Cela se traduit par l'inégalité $(p-1)|t| \geq |t|_e$,

Donc

$$\begin{aligned} |t_e| R_t(\rho) R_e(\rho)^{-1} &= |t_e| \rho^{|t|} \prod_{i=2}^{q} \pi_{e_i}^{|t|_{e_i}} R_e(\rho)^{|t|_e - 1} \\ &< (p-1)\, |t| \rho^{|t|} \tau^{(p-1)|t|-1} \end{aligned}$$

Or pour tous $p \geqslant 2$ et $x \geqslant 1$, on a

$$(p-1)\, x \left(\frac{(p-1)^{p-1}}{p^p} \right)^x \left(\frac{p}{p-1} \right)^{(p-1)x-1} = x \frac{(p-1)^2}{p^{x+1}} < 1,$$

donc on a bien $|t_e| R_t(\rho) R_e(\rho)^{-1} \leq 1$ avec parfois inégalité stricte. Donc

$$\sum_{t \in \mathcal{I}} |t_e| R_t^p(\rho) R_e(\rho)^{-1} < \sum_{t \in \mathcal{I}} R_t^{p-1}(\rho) < \left(\sum_{t \in \mathcal{I}} R_t(\rho) \right)^{p-1}$$

et

$$\frac{\partial F}{\partial y}(\rho, R_e(\rho)) = 1 - p\rho \sum_{t \in \mathcal{I}} |t_e| R_t^p(\rho) R_e(\rho)^{-1} > 1 - p\rho \left(\sum_{t \in \mathcal{I}} R_t(\rho) \right)^{p-1} = 1 - m\rho R(\rho)^{p-1} = 0.$$

□

À ce titre, nous utiliserons le lemme suivant, qui caractérise un ensemble de bornes supérieures :

Lemme 4.7 *Soient* $f(y) := 1 - \pi(e) + y^p - y$, *et* $x > 0$. *Si* $f(\tau - x) > 0$, *alors* $R_e(\rho) < x$.

Démonstration. Soit $\mathcal{R}_{\bar{e}}$ la famille des arbres de $\mathcal{R}$ ne comportant pas l'étiquette e. On peut injecter $\mathcal{R}_{\bar{e}}$ dans un sous-ensemble de $\mathcal{R}$ dont aucun arbre ne se réduit en e. Cette injection ϕ se définit de la manière suivante :

$$\begin{array}{lll} \phi : \mathcal{R}_{\bar{e}} & \rightarrow & \mathcal{R} \setminus \mathcal{R}_e \\ t \notin \mathcal{R}_e & \mapsto & t \\ t \in \mathcal{R}_e & \mapsto & t' \end{array}$$

Si $t \in \mathcal{R}_{\bar{e}}$, deux cas de figure se présentent :

a) t possède un nœud interne de profondeur maximale tel que tous ses fils sont identiques, égaux à e'. Alors t' est l'arbre obtenu à partir de t en transformant le premier fils en e.

b) t' possède un nœud interne de profondeur maximale avec p fils qui sont des nœuds internes dont pendent (éventuellement via d'autres nœuds internes) une même succession d'étiquettes. Du premier fils (par exemple), on remplace la deuxième étiquette par un e.

Il est clair que ϕ est une injection stricte de $\mathcal{R}_{\bar{e}}$ dans $\mathcal{R} \setminus \mathcal{R}_e$, puisque dans le cas b) par exemple, on aurait pu choisir de remplacer la deuxième étiquette du deuxième fils.

Donc $(R_{\bar{e}} + R_e)(\rho) < R(\rho)$.

$$\begin{array}{rcl} R(\rho) & = & 1 + \rho R^p(\rho) = \tau \\ R_{\bar{e}}(\rho) & = & 1 - \pi(e) + \rho R^p_{\bar{e}}(\rho) \end{array}$$

Si $x > \tau$, $R_e(\rho) < \tau < x$.

Si $x < \tau$, on aura $R_e(\rho) < x < \tau$ si

$$\begin{array}{rcl} R(\rho) - R_{\bar{e}}(\rho) & < & x \\ R(\rho) - x = \tau - x & < & R_{\bar{e}}(\rho) \\ \tau - x & < & 1 - \pi(e) + \rho R^p_{\bar{e}}(\rho) \end{array}$$

x	0	$R_{\bar{e}}(\rho)$	τ		$+\infty$
$f'(x)$		$-$	0	$+$	$+\infty$
$f(x)$	$1-\pi(e)$	$\searrow$	$-\pi(e)$	$\nearrow$	$+\infty$

On sait que $\tau - x < \tau$, donc si $f(\tau - x) > 0$, cela signifie que $\tau - x < R_{\bar{e}}(\rho)$. □

Lemme 4.8 *Pour* $p = 2$, $R_e(\rho) < 2\sqrt{\pi(e)}$.

Démonstration. $g(\tau - x) > 0$ s'écrit

$$\begin{array}{rcl} \tau - x & < & 1 - \pi(e) + \rho(\tau - x)^p, \\ 1 + \rho\tau^p - x & < & 1 - \pi(e) + \rho(\tau - x)^p, \\ \dfrac{\pi(e) - x}{\rho} + \tau^p - (\tau - x)^p & < & 0. \end{array}$$

Pour $p = 2$, il vient $4\pi(e) < x^2$. En fait on a une équation du second degré et le résultat exact $R_{\bar{e}}(\rho) = \tau - 2\sqrt{\pi(e)}$. Donc $R_e(\rho) < R(\rho) - R_{\bar{e}}(\rho) = 2\sqrt{\pi(e)}$. □

Lemme 4.9 *Il existe $c > 0$, $z_0 \in]\rho, \rho_I[$, $\gamma < \sqrt{\rho_I}$ tels que pour tout $t \in \mathcal{I}$,*

$$R_t(z_0) < \sqrt{\pi_t} c \gamma^{|t|}. \quad .$$

Démonstration. De $\rho = 1/4$ on déduit $\sqrt{\rho} R_e(\rho) < \sqrt{\pi(e)}$.

Pour tout t, nous avons $R_t(z) = \prod_{i=2}^{q} \pi_i^{|t|_i} z^{|t|-|t|_e} R_e(z)^{|t|_e}$, si bien que

$$R_e(z) = \pi_e + \sum_{t \in \mathcal{I}} \prod_{i=2}^{q} \pi_i^{|t|_i} z^{|t|-|t|_e} R_e(z)^{|t|_e}.$$

Une application du théorème des fonctions implicites permet de montrer que $R_e(z)$ est encore analytique au voisinage de ρ, si bien R_e est analytique au-delà de ρ. Soit $z_0 \in]\rho, \rho_I[$ tel qu'on ait encore

$$\sqrt{z_0} R_e(z_0) < \sqrt{\pi(e)}.$$

Soient alors $c = z_0^{-1/2}$, et $\gamma = \sqrt{z_0}$. On a $R_e(z_0) < c\pi_e \gamma^0$. Remarquons $c > 1$, donc pour toute étiquette $a \neq e$, $R_a(z_0) = \pi_a < c\pi_a \gamma^0$.

Par récurrence, on obtient que pour $t = \circ(u, v)$ de taille ≥ 1,

$$R_t(z_0) = z_0 (R_u R_v)(z_0) < \pi_u \pi_v z_0 c^2 \gamma^{|u|+|v|} = \pi_t c (z_0 c) \gamma^{|t|-1} = \pi_t c \gamma^{|t|},$$

puisque $z_0 c = \sqrt{z_0} = \gamma$. □

La preuve d'une inégalité du type $R_t(z_0) < \pi_t c \gamma^{|t|}$ avec $\gamma^p < \rho_I$ peut encore se faire de manière similaire pour $p = 3$ en utilisant la relation de récurrence $R_t(z) = z \prod_{j=1}^{p} R_{t_j}(z)$, mais plus pour $p \geqslant 4$, où il sera nécessaire d'utiliser une relation du type

$$\prod_{i=2}^{q} \pi_i^{|t|_i} z^{|t|-|t|_e} R_e(z)^{|t|_e},$$

où $|t|_e$ est le nombre de fois où le symbole e apparaît dans l'arbre t. Ce deuxième schéma de preuve marchera pour $p \geqslant 3$ (il ne marcherait pas pour $p = 2$, comme le montreront les calculs).

Pour établir cette inégalité dans le cas $p \geqslant 3$, nous introduisons afin de simplifier les calculs la série génératrice comptant les arbres par le nombre total de nœuds, $M(z) = z + zM^p(z)$, avec p l'arité de l'arbre, et M_t la série génératrice des arbres se réduisant en t comptés par nombre total de nœuds. Pour tout arbre t, on note $\| t \|$ son nombre total de nœuds. Nous montrons le lemme suivant :

Lemme 4.10 *Soient ρ' la singularité dominante de M, et ρ'_I la singularité dominante de la série génératrice des arbres irréductibles comptés par nombre total de nœuds. Alors il existe $\varepsilon' > 0$ tel que pour tout arbre t, M_t possède un rayon de convergence au moins égal à $\rho' + \varepsilon'$ et*

$$M_t(\rho' + \varepsilon') \leqslant \pi_t^{1/p} \rho'^{\|t\|/2}_I.$$

Démonstration. Soit $\tau' = M(\rho')$. Nous avons

$$\rho' = \frac{p-1}{p} \sqrt[p]{\frac{1}{p-1}}; \qquad \tau' = \sqrt[p]{\frac{1}{p-1}} = \frac{p}{p-1} \rho' < 1.$$

Montrons que M_e possède un rayon de convergence strictement supérieur à ρ'. Pour tout t, nous avons $M_t(z) = \prod_{i=2}^{q} \pi_i^{|t|_i} z^{\|t\|-|t|_e} M_e(z)^{|t|_e}$, si bien que

$$M_e(z) = \pi_e z + \sum_{t \in \mathcal{I}} \prod_{i=2}^{q} \pi_i^{|t|_i} z^{\|t\|-|t|_e} M_e(z)^{|t|_e}.$$

Une application du théorème des fonctions implicites permet de montrer que $M_e(z)$ est encore analytique au voisinage de ρ', si bien que ρ' est strictement inférieur à son rayon de convergence.

Il est clair que $M_e(\rho') < \tau'$. Soit $\varepsilon' > 0$ tel que $\rho' + \varepsilon' < \rho'_I$ et que $M_e(\rho' + \varepsilon')$ soit convergent, avec toujours $M_e(\rho' + \varepsilon') < \tau'$. L'expression $M_t(z) = \prod_{i=2}^{q} \pi_i^{|t|_i} z^{\|t\|-|t|_e} M_e(z)^{|t|_e}$ montre que les M_t ont même rayon de convergence que M_e.

On aura

$$\begin{aligned} M_t(\rho'+\varepsilon') = \prod_{i=2}^{q} \pi_i^{|t|_i} (\rho'+\varepsilon')^{\|t\|-|t|_e} M_e(\rho'+\varepsilon')^{|t|_e} &\leqslant \pi_t^{1/p} \sqrt{\rho'_{\mathcal{I}}}^{\|t\|} = \prod_{i=1}^{q} \pi_i^{|t|_i/p} \sqrt{\rho'_{\mathcal{I}}}^{\|t\|} \\ (\rho'+\varepsilon')^{\|t\|-|t|_e} M_e(\rho'+\varepsilon')^{|t|_e} &\leqslant \prod_{i=1}^{q} \pi_i^{|t|_i/p-1} \pi_e^{|t|_e} \sqrt{\rho'_{\mathcal{I}}}^{\|t\|} \end{aligned}$$

si

$$\begin{aligned} \rho'^{\|t\|-|t|_e}_{\mathcal{I}} M_e(\rho'+\varepsilon')^{|t|_e} &\leqslant \prod_{i=1}^{q} \pi_i^{|t|_i/p-|t|_i} \pi_e^{|t|_e} \sqrt{\rho'_{\mathcal{I}}}^{\|t\|} \\ M_e(\rho'+\varepsilon')^{|t|_e} &\leqslant \prod_{i=1}^{q} \pi_i^{|t|_i/p-|t|_i} \pi_e^{|t|_e} \rho'^{|t|_e-\|t\|/2}_{\mathcal{I}} \end{aligned}$$

Or pour tout $t \neq e \in \mathcal{I}$,

$$|t|_e \leqslant \frac{p-1}{p} \|t\|, \quad (\text{ et } \rho'_{\mathcal{I}} < 1).$$

En effet, si $\|t\|$ est le nombre total de nœuds d'un arbre p-aire et $|t|$ son nombre de nœuds internes, alors $\|t\| = p|t| + 1$, ce qui entraîne $|t| < \|t\|/p$. Dans un arbre irréductible, tout nœud interne a au moins un fils différent de e. Cela se traduit par l'inégalité $(p-1)|t| \geq |t|_e$, puis en utilisant l'inégalité précédente $|t|_e < \frac{p-1}{p} \|t\|$. Il suffit donc

$$\begin{aligned} M_e(\rho'+\varepsilon')^{|t|_e} &\leqslant \prod_{i=1}^{q} \pi_i^{|t|_i/p-|t|_i} \pi_e^{|t|_e} \rho'^{|t|_e-\frac{p}{p-1}|t|_e/2}_{\mathcal{I}} = \pi_t^{\left(\frac{1}{p}-1\right)\frac{p}{p-1}|t|_e} \pi_e^{|t|_e} \rho'^{|t|_e-\frac{p}{p-1}|t|_e/2}_{\mathcal{I}} \\ M_e(\rho'+\varepsilon') &\leqslant \pi_e^{\left(\frac{1}{p}-1\right)\frac{p}{p-1}} \pi_e \prod_{i=2}^{q} \pi_i^{|t|_i/p-|t|_i} \rho'^{1-\frac{p}{p-1}/2}_{\mathcal{I}} = \prod_{i=2}^{q} \pi_i^{\left(\frac{1}{p}-1\right)|t|_i} \rho'^{\frac{p-2}{2(p-1)}}_{\mathcal{I}} \end{aligned}$$

Or

$$\prod_{i=2}^{q} \pi_i^{\left(\frac{1}{p}-1\right)|t|_i} > 1,$$

donc il suffit

$$M_e(\rho'+\varepsilon') \leqslant \rho'^{\frac{p-2}{2(p-1)}}_{\mathcal{I}}.$$

En fait, on montre $\tau' < \rho'^{\frac{p-2}{2(p-1)}}$: ce sera vrai si

$$\begin{array}{rcl} \left(\frac{1}{p-1}\right)^{\frac{1}{p}} & < & \left(\frac{p-1}{p}\left(\frac{1}{p-1}\right)^{\frac{1}{p}}\right)^{\frac{p-2}{2(p-1)}} \\ \Longleftarrow \quad p^{\frac{p-2}{2(p-1)}} & < & (p-1)^{\left(1-\frac{1}{p}\right)\frac{p-2}{2(p-1)}+\frac{1}{p}} = (p-1)^{\frac{p-2}{p}+\frac{1}{p}} = (p-1)^{\frac{p-1}{p}} \\ \Longleftarrow \quad p^{\frac{1}{2}-\frac{1}{2(p-1)}} & < & (p-1)^{1-\frac{1}{p}} \\ \Longleftarrow \quad 1 & < & (p-1)^{1-\frac{1}{p}} p^{\frac{1}{2(p-1)}-\frac{1}{2}} \\ \Longleftarrow \quad 0 & < & \left(1-\frac{1}{p}\right)\log(p-1) + \frac{1}{2}\left(\frac{1}{p-1}-1\right)\log p. \end{array}$$

Le membre de droite vaut $0,187... > 0$ en $p = 3$ (et nul en $p = 2$), et sa dérivée

$$\frac{1}{p} + \frac{1}{p^2}\log(p-1) + \frac{1}{2p(p-1)} - \frac{1}{2p} = \frac{1}{2p} + \frac{1}{p^2}\log(p-1) + \frac{1}{2p(p-1)}$$

est strictement positive, donc c'est une fonction croissante en p, et l'inégalité est vérifiée pour $p \geqslant 3$. Donc on a bien $M_e(\rho') < \tau' < \rho'^{\frac{p-2}{2(p-1)}} < \rho'^{\frac{p-2}{2(p-1)}}_{\mathcal{I}}$. □

Lemme 4.11 *Soit $p \geqslant 3$. Il existe $c > 0$, $z_0 \in]\rho, \rho_I[$, $\gamma^p < \rho_I$ tels que pour tout $t \in \mathcal{I}$, $R_t(z_0) < c\pi_t^{1/p}\gamma^{|t|}$.*

Démonstration. Fixons $p \geqslant 3$. Soit $\varepsilon' > 0$ comme dans le lemme précédent, tel que pour tout t,

$$M_t(\rho' + \varepsilon') < \pi_t^{1/p}\rho_I'^{\|t\|/2}.$$

Posons $z_0 = (\rho' + \varepsilon')^p$. Étant donné $M_t(z) = zR_t(z^p)$, $\rho = \rho'^p$ et $\rho_I = \rho_I'^p$, il vient :

$$\begin{array}{rcl} (\rho' + \varepsilon')R_t((\rho' + \varepsilon')^p) & < & \pi_t^{1/p}\rho_I'^{\|t\|/2} \\ z_0^{\frac{1}{p}} R_t(z_0) & < & \pi_t^{1/p}\rho_I'^{\frac{p|t|+1}{2}} \\ R_t(z_0) & < & z_0^{-\frac{1}{p}}\rho_I^{\frac{1}{2p}}\pi_t^{1/p}\rho_I^{\frac{|t|}{2}} \end{array}$$

Il suffit de poser $c = z_0^{-\frac{1}{p}}\rho_I^{\frac{1}{2p}}$ et $\gamma = \rho_I^{1/2}$ pour avoir le résultat voulu, puisque

$$\gamma^p = \rho_I^{\frac{p}{2}} = \rho_I^{\frac{p-2}{2}} \cdot \rho_I,$$

avec $\rho_I^{(p-2)/2} < 1$ dès que $p > 2$. □

Lemme 4.12 *Soit $g(z,y) = \sum_{t\in\mathcal{I}} \left(y^{m|t|+1} - 1\right) R_t^m(z)$ la fonction telle que $H(z,y) = 1 + zH(z,y)^m - zg(z,y)$. Il existe $z_0 > \rho_R$ et $y_0 > 1$ tels que $g(z,y)$ est holomorphe en les deux variables z et y dans le domaine $|z| < z_0$, $|y| < y_0$.*

Démonstration. Soient c, z_0 et γ comme dans les propositions 4.9 ou 4.11 suivant $p = 2$ ou $p \geqslant 3$. Soit $y_0 \in \left]1, \sqrt{\rho_I\gamma^{-p}}\right[$. Alors pour $|z| < z_0$, $|y| < y_0$, les relations suivantes sont satisfaites :

$$\sum_{t\in\mathcal{I}} \left(y^{p|t|+1}\right) R_t^p(z) \leqslant \sum_{t\in\mathcal{I}} \left(y_0^{p|t|+1}\right) R_t^p z_0) \leqslant \sum_{n\geq 0} I_n \left(y_0^{pn+1}\right) c^p\gamma^{pn},$$

$$\sum_{t\in\mathcal{I}} R_t^p(z) \leqslant \sum_{t\in\mathcal{I}} R_t^p(z_0) \leqslant \sum_{n\geq 0} I_n c^p \gamma^{pn}.$$

De $y_0^p\gamma^p < \rho_I$ et $\gamma^p < \rho_I$, nous déduisons que les deux séries numériques

$$\sum_{t\in\mathcal{I}} y^{p|t|+1} R_t^p(z) \qquad \text{et} \qquad \sum_{t\in\mathcal{I}} R_t^p(z)$$

convergent. Comme les deux séries ont des coefficients positifs en $z^n y^k$, elles sont normalement convergentes, et donc holomorphes en z et y. Leur différence $g(z,y)$ est par conséquent aussi holomorphe en z et y. □

Comme c'était le cas dans la section 4.2, nous pouvons montrer que H suit un schéma algébrique :

Proposition 4.3 *Il existe des réels $z_1 > \rho$, $y_1 > 1$, et des fonctions analytiques A, B et C telles que, dans le domaine*

$$\mathcal{D} = \{(z,y) \mid |z| < z_1,\ |y| < y_1\},$$

H admet la représentation $H(z,y) = A(z,y) + B(z,y)\sqrt{C(z,y)}$.

Démonstration. Nous procédons comme dans la preuve de la Proposition 4.1, en considérant la function $(z,y,x) \mapsto x - 1 - yzx^p + zg(z,y)$. □

4.3.3 Loi limite

Nous sommes maintenant capables de déterminer la distribution limite de la taille des arbres réduits.

Théorème 4.2 *On suppose la condition de variabilité vérifiée (i.e. $\varliminf_{n\to\infty} \frac{\sigma_n}{n} > 0$). Alors la taille des arbres p-aires réduits admet une distribution limite gaussienne, et la moyenne μ_n et la déviation standard σ_n sont asymptotiquement linéaires en n, avec*

$$\lim \frac{\mu_n}{n} = \mu \text{ et } \lim \frac{\sigma_n}{n} = \overline{v},$$

où $\mu = 1 - \rho(p-1)g_y(\rho,1)$ et

$$\begin{aligned}\overline{v} = &\left[\rho(p-1) - (p-1)(p-2)\rho^2\right] g_y(\rho,1) - 3(p-1)^2\rho^2 g_y^2(\rho,1)\\ &+ 2\rho^2 p - 1 g_{yz}(\rho,1) - 2\rho^3 (p-1)^2 g_y(\rho,1) g_{yz}(\rho,1) - (m-1)\rho g_{y^2}(\rho,1).\end{aligned}$$

Démonstration. Elle est identique à celle du Théorème 4.1. □

Des calculs numériques pour $p = q = 2$ et $\pi_e = \frac{1}{2}$ donnent les valeurs $\mu = 0.8162...$ et $\overline{v} = 0.2469....$ Le taux de réduction moyen n'est pas beaucoup plus petit que le taux de simplification moyen.

4.4 Règles d'absorption

On considère des expressions booléennes, représentées par des arbres d'expressions qui sont des arbres binaires complets, soit réduits à une feuille, à savoir V, F ou un littéral, soit formé à la racine d'un opérateur $\vee$ ou $\wedge$, et de deux sous-arbres d'expression booléenne.

Nous allons voir qu'on a toujours une distribution limite discrète, avec un jeu de réductions plus restreint.

On introduit les règles de réduction suivante :

Absorption (4 règles)

$$\begin{aligned} f \wedge F &\to F \\ F \wedge f &\to F \\ f \vee V &\to V \\ V \vee f &\to V \end{aligned}$$

Élément neutre (4 règles)

$$\begin{aligned} f \wedge V &\to f \\ V \wedge f &\to f \\ f \vee F &\to f \\ F \vee f &\to f \end{aligned}$$

Par la suite on notera $\odot$ et $\oplus$ les opérateurs $\wedge$ et $\vee$, pour souligner que nous nous trouvons dans une algèbre de Boole.

On se place dans le modèle suivant : les opérateurs $\odot$ et $\oplus$ sont équiprobables. On tire V et F avec des probabilités chacune égale à $p/2$, et on tire une variable générique X avec probabilité $q = 1-p$. On associe à chaque arbre son poids, qui est le produit des probabilités de ses opérandes et de ses opérateurs.

Notons qu'étant donné les règles de réductions considérées, on n'a pas besoin de distinguer les variables les unes des autres, si bien qu'on notera X n'importe laquelle de ces variables.

Nous appellerons taille d'une expression le nombre de ses opérandes, ce qui revient au nombre de feuilles des arbres correspondants.

Notre modèle est différent de celui qui est étudié [31] puisque nous ne considérons pas toutes les règles de réduction. Par ailleurs nous étudions la taille limite des arbres réduits, et pas la probabilité qu'un arbre représente une fonction booléenne, classe d'équivalence d'expressions booléennes.

4.4.1 Séries génératrices univariées

On s'intéresse d'abord aux séries génératrices des arbres se réduisant en un arbre donné.

Proposition 4.4 *Soit, pour tout arbre t, S_t la série génératrice des arbres se réduisant en t. Soit B la série génératrice des arbres de Catalan. Alors*

$$\begin{aligned} S_V &= \frac{p}{2}B, \\ S_F &= S_V, \end{aligned}$$

et pour tout $f = f_1 \odot f_2$ ou $f = f_1 \oplus f_2$,

$$S_f = \frac{1}{2}S_{f_1}S_{f_2} + 2S_V S_f.$$

Démonstration. Les séries génératrices s'écrivent

$$\begin{aligned} S_V &= \frac{p}{2}z + [f \oplus V]\frac{1}{2}(B - S_V)S_V + [V \oplus f]\frac{1}{2}S_V(B - S_V) \\ &\quad + [V \oplus V]\frac{1}{2}S_V^2 + [V \odot V]\frac{1}{2}S_V^2 \\ &= \frac{p}{2}z + BS_V = \frac{pz/2}{1-B} = \frac{pz}{1+\sqrt{1-4z}} \\ &= \frac{pz(1-\sqrt{1-4z})}{4z} = \frac{p}{2}B, \\ S_F &= S_V. \end{aligned}$$

On a 4 règles d'élément neutre, chacune avec poids 1/2, ce qui se somme à 2, d'où

$$\begin{aligned} S_X &= qz + 2S_V S_X = qz + pBS_X = \frac{qz}{1-pB} = \frac{qz}{1-p\left(\frac{1-\sqrt{1-4z}}{2}\right)} \\ &= \frac{qz\left(1-\frac{p}{2}-\frac{p}{2}\sqrt{1-4z}\right)}{\left(1-\frac{p}{2}\right)^2 - \frac{p^2}{4} + p^2 z} \\ &= \frac{qz\left(1-\frac{p}{2}-\frac{p}{2}\sqrt{1-4z}\right)}{1-p+p^2 z} \\ &= \frac{qz\,(q+pB)}{q+p^2 z}, \end{aligned}$$

$$S_f = \frac{1}{2} S_{f_1} S_{f_2} + 2S_V S_f.$$

□

Notons que la singularité $-q/p^2$ qui semble apparaître dans la dernière expression de S_X est artificielle, et qu'elle apparaît parce que nous avons voulu avoir un dénominateur sans radical, et multiplié numérateur et dénominateur de la première fraction par la quantité conjuguée de $1-pB$.

On introduit la notation $\square$ pour une étiquette V, F ou X. En particulier, $S_\square = S_V + S_F + S_X$.

4.4.2 Série génératrice bivariée

Nous introduisons maintenant la série génératrice bivariée énumérant les arbres à la fois par taille (variable z) et par taille de l'arbre en lequel ils se réduisent (variable y).

Proposition 4.5 *Soient $\mathcal{I}$ l'ensemble des arbres irréductibles et $H(z,y)$ la fonction génératrice bivariée des arbres d'expressions booléennes énumérés par taille et taille d'arbre réduit. Alors*

$$H(z,y) = yS_\square + (H - 2yS_V)^2 + 2S_V\,(H - 2yS_\square)\,.$$

Corollaire 4.1 *H admet pour expression*

$$H(z,y) = \frac{1 - 2(1-y)S_V - \sqrt{(1-2(1-y)S_V)^2 - 4\,(yS_\square - 4S_V S_\square + 4y^2 S_V{}^2)}}{2}$$

Démonstration. Introduisons les sous-ensembles de $\mathcal{I}$ suivants :

$$\mathcal{I}^* = \mathcal{I} \setminus \{V, F\}\ ; \quad \mathcal{I}^{**} = \mathcal{I} \setminus \{V, F, X\}.$$

On a alors

$$\begin{aligned}
H(z,y) &= \sum_{t\in\mathcal{B}} y^{|s(t)|}z^{|t|} = \sum_{t\in\mathcal{I}} y^{|t|}S_t(z)\\
&= yS_\square + \sum_{t\in\mathcal{I}^{**}} y^{|t|}S_t(z)\\
&= yS_\square + \sum_{t_1,t_2\in\mathcal{I}^*} y^{|t_1|+|t_2|}\left(S_{t_1\oplus t_2}(z) + S_{t_1\odot t_2}(z)\right)\\
&= yS_\square + \sum_{t_1,t_2\in\mathcal{I}^*} y^{|t_1|+|t_2|}\left(\frac{1}{2}S_{t_1}S_{t_2}(z) + 2S_VS_{t_1\oplus t_2} + \frac{1}{2}S_{t_1}S_{t_2}(z) + 2S_VS_{t_1\odot t_2}\right)\\
&= yS_\square + \sum_{t_1,t_2\in\mathcal{I}^*} y^{|t_1|+|t_2|}S_{t_1}S_{t_2} + 2S_V\sum_{t_1,t_2\in\mathcal{I}^*;\ t=o(t_1,t_2)} y^{|t|}\left(S_{t_1\oplus t_2} + S_{t_1\odot t_2}\right)\\
&= yS_\square + (H-2yS_V)^2 + 2S_V\left(H-2yS_\square\right).
\end{aligned}$$

□

4.4.3 Distribution limite

La proposition précédente permet d'avoir une expression exacte de $H(z,y)$. Pour en extraire la distribution limite, nous allons en fait calculer exactement les séries génératrices des arbres de paramètre k (taille de l'arbre réduit), puis par analyse de singularité trouver la probabilité qu'un arbre, dont la taille tend vers l'infini, se réduise en un arbre de taille k.

Théorème 4.3 *La taille des arbres réduits selon les règles d'absorption ou de multiplication par un élément neutre admet une limite discrète, donnée par la série génératrice*

$$g(y) = py + pf\left(\frac{qy}{(1+q)^2}\right),$$

où on a posé

$$f(x) = \frac{2x}{\sqrt{1-4x}} - \frac{1-\sqrt{1-4x}}{2}.$$

Démonstration. Posons $K(z,y) = H(z,y) - 2yS_V(z)$. Il vient

$$\begin{aligned}
K &= (1-2S_V)yS_X + 2S_VK + K^2\\
&= (1-2S_V)\left(\frac{1-\sqrt{1-4y\frac{S_X}{1-2S_V}}}{2}\right).
\end{aligned}$$

Pour $k \geqslant 1$ on a, en utilisant les expressions précédemment trouvées

$$S_X(z) = \frac{qz}{1-pB(z)} \text{ et } S_V = \frac{p}{2}B,$$

$$\begin{aligned}
[y^k]K &= (1-2S_V)\left(\frac{S_X}{1-2S_V}\right)^k C_{k-1} \\
&= (qz)^k(1-pB)^{1-2k}C_{k-1} \\
&= \frac{(qz)^k C_{k-1}}{(1-pB)^{2k-1}} \\
&= (qz)^k C_{k-1}\left(\frac{q+pB}{q+p^2z}\right)^{2k-1} \\
&= (qz)^k C_{k-1}(q+p^2z)^{-(2k-1)}\left(q+\frac{p}{2}\left(1-\sqrt{1-4z}\right)\right)^{2k-1} \\
&= (qz)^k C_{k-1}(2q+2p^2z)^{-(2k-1)}\left(2q+p\left(1-\sqrt{1-4z}\right)\right)^{2k-1} \\
&= (qz)^k C_{k-1}(2q+2p^2z)^{-(2k-1)}\left(2q+p-p\sqrt{1-4z}\right)^{2k-1} \\
&= (qz)^k C_{k-1}(2q+2p^2z)^{-(2k-1)}\left((2q+p)-p\sqrt{1-4z}\right)^{2k-1}.
\end{aligned}$$

Donc

$$\begin{aligned}
[z^n y^k]K &\underset{n\to\infty}{\sim} [z^n](qz)^k C_{k-1}(2q+2p^2z)^{-(2k-1)}(2k-1)(2q+p)^{2k-2}(-p\sqrt{1-4z}) \\
&\underset{n\to\infty}{\sim} 2p(q\tfrac{1}{4})^k C_{k-1}(2q+2p^2\tfrac{1}{4})^{-(2k-1)}(2k-1)(2q+p)^{2k-2}C_{n-1}.
\end{aligned}$$

D'autre part

$$S_\square = S_X + 2S_V = \frac{qz}{1-pB} + pB = \frac{(q+pB)qz}{q+p^2z} + pB,$$

d'où

$$[z^n]S_\square \underset{n\to\infty}{\sim} \left(\frac{pq/4}{q+p^2/4}+p\right)C_{n-1}.$$

Donc pour tout k on a

$$\lim_{n\to\infty}\frac{[z^n y^k]H(z,y)}{[z^n]H(z,1)} = p_k,$$

avec $g(y)=\sum_k p_k y^k$ d'expression

$$\begin{aligned}
g(y) &= \left(\frac{pq/4}{q+p^2/4}+p\right)y + \sum_{k\geqslant 2}\frac{(q/4)^k C_{k-1}}{(2q+2p^2/4)}2p(2k-1)(2q+p)^{2k-2}y^k \\
&= py + \frac{2q+2p^2/4}{(2q+p)^2}\sum_{k\geqslant 1}(2k-1)C_{k-1}\left(\frac{q/4(2q+p)^2y}{(2q+2p^2/4)^2}\right)^k.
\end{aligned}$$

Si on pose

$$f(x) = \frac{2x}{\sqrt{1-4x}} - \frac{1-\sqrt{1-4x}}{2},$$

alors

$$g(y) = py + 2p\frac{2q+2p^2/4}{(2q+p)^2}f\left(\frac{q/4(2q+p)^2y}{(2q+2p^2/4)^2}\right) = py + pf\left(\frac{qy}{(1+q)^2}\right).$$

On a toujours $\frac{q}{(1+q)^2} < \frac{1}{4}$ car $0 < (1-q)^2$ donc f est analytique au-delà de 1, de même pour g.

On vérifie $g(1) = 1$ car $f(x) = \frac{1}{2\sqrt{1-4x}} - \frac{1}{2}$ et

$$\begin{aligned} g(1) &= p + p\frac{1}{2\sqrt{1-4\frac{q}{(1+q)^2}}} - \frac{1}{2} \\ &= p + p\frac{1}{2\left(\frac{1-q}{1+q}\right)} - \frac{1}{2} \\ &= \frac{2p(1+q)+p(1-q)}{2(1-q)} - \frac{1}{2} \\ &= \frac{3p}{2p} - \frac{1}{2} \\ &= 1. \end{aligned}$$

On en déduit que g est bien une série génératrice de probabilité, et on a une limite discrète donnée par g. □

Corollaire 4.2 *La moyenne de la taille limite des arbres réduits dans le modèle précédent vaut*

$$\frac{p^3 - 2p^2 + 6p - 2}{p^2}.$$

Démonstration. Il suffit de calculer $g'(1)$.

$$\begin{aligned} g'(y) &= p - p\frac{q}{(2-p)^2}\frac{2}{\left(1-\frac{4qy}{(2-p)^2}\right)^{3/2}} \\ g'(1) &= p - \frac{pq}{(1+q)^2}\frac{2}{\left(\frac{1-q}{1+q}\right)^3} \\ &= p - \frac{2pq(1+q)}{(1-q)^3} \\ &= p - 2\frac{(1-p)(2-p)}{p^2} \\ &= \frac{p^3 - 2p^2 + 6p - 2}{p^2}. \end{aligned}$$

□.

4.5 Conclusion et questions ouvertes

Nous avons prouvé que la taille des arbres simplifiés (loi idempotente) et réduits (loi nilpotente) sont asymptotiquement gaussiennes, alors que seules la moyenne et la variance étaient connues avant. Notre schéma de travail permet aussi d'affirmer, dans le cadre du théorème des quasi puissances, que la vitesse de convergence vers une limite gaussienne est en $O(1/\sqrt{n})$ grâce aux inégalités de Berry-Essén.

Pour les lois idempotente et nilpotente, on voit que la simplification ou la réduction n'apporte pas un gain en place probant. En revanche, la présence de lois de type absorbant mène à des lois limites discrètes, ce qui rend la réduction très intéressante.

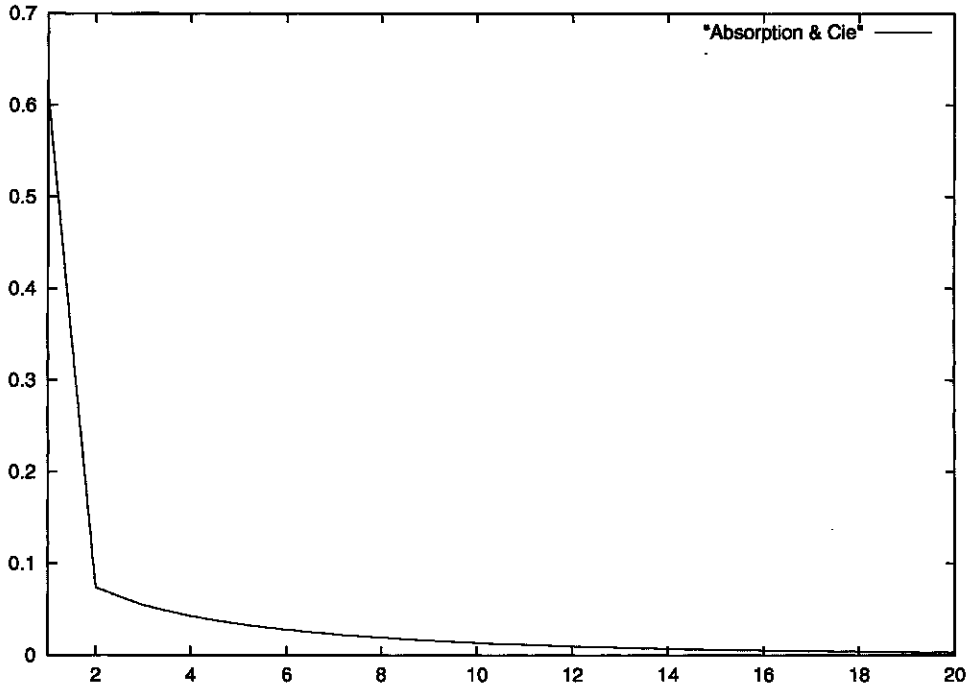


FIG. 4.1 – Distribution limite de le réduction booléenne avec $p = \frac{1}{2}$.

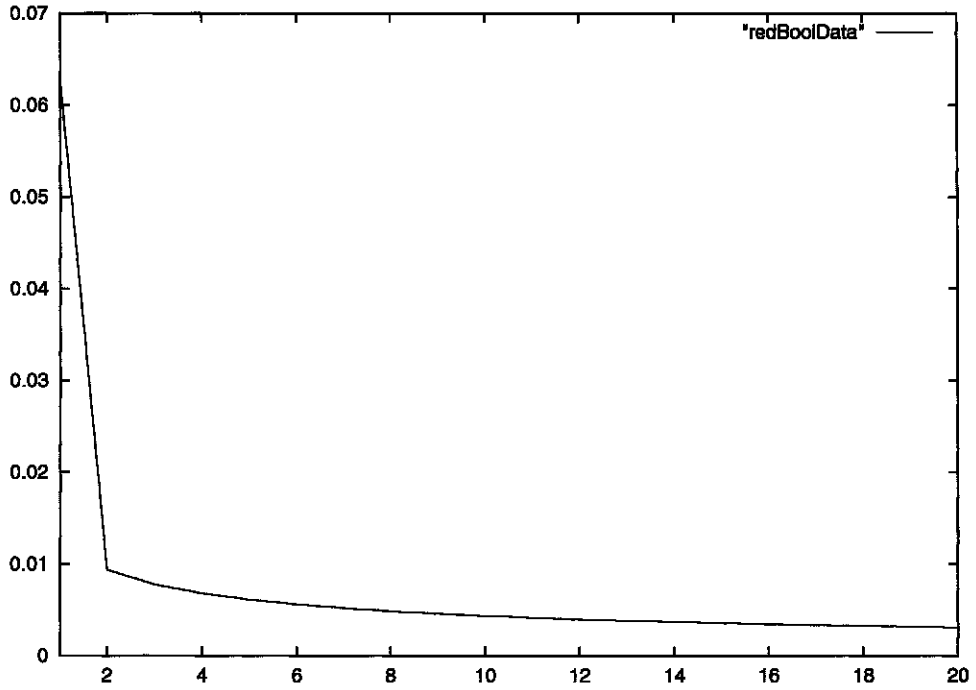


FIG. 4.2 – Distribution limite de le réduction booléenne avec $p = \frac{1}{20}$.

Un développement naturel consisterait à étendre cette étude à une plus large famille d'arbres, et à mélanger de manière plus sélective les différents types d'opérateurs.

Un défi aussi difficile et intéressant consisterait à étudier les distributions des limites des coûts des algorithmes de simplification et de réduction. Leurs moyennes sont connues (pour les lois idempotente et nilpotente) et sont linéaires en n, mais leurs variances n'ont pas encore été calculées. Leurs distributions en sont d'autant plus inconnues.

Chapitre 5

Énergie de structures secondaires d'ARN

Les molécules d'ARN jouent un rôle fondamental dans la transmission du message génétique contenu dans l'ADN aux protéines chargées des fonctions cellulaires. Leur forme de repliement est liée à leur fonction biologique, si bien qu'elle revêt une grande importance pour les biologistes. Les structures secondaires, version bidimensionnelle de ce repliement, capturent déjà une grande partie de l'information contenue dans cette forme de repliement.

Depuis que M.S. Waterman [209, 208] a formalisé les structures secondaires d'ARN, celles-ci font l'objet d'une étude intensive de la part des informaticiens [24, 55, 56, 92, 97, 117, 184, 197, 196, 216], en particulier autour de la prédiction de la forme des structures secondaires à partir d'un mot d'ARN, qui se traduit souvent par la minimisation de l'énergie de la molécule d'ARN calculée à partir des structures secondaires possibles.

Très vite sont apparues des séries génératrices pour énumérer les molécules d'ARN [106], et Markus Nebel, tout récemment, donne une approche plus fine de ces molécules pour étudier la distribution de la présence de conformations particulières type boucle, épingle, etc. (voir la définition de ces conformations particulières plus bas) [163, 166, 167, 168].

Dans ce chapitre, nous étudions la distribution limite de l'énergie de structures secondaires d'ARN tirées de manière complètement aléatoire, dans un modèle simplifié à base de grammaires S-attribuées [130, 136, 137, 135], qui généralisent les grammaires attribuées. Dans un modèle encore plus simple, Xiong et Waterman [212] montrent l'existence d'une distribution limite de l'énergie optimale à l'aide d'une généralisation du théorème de Kingman [126] (affirmant la convergence de suites sous-additives vers sa limite inférieure), appliquée à la surmartingale qu'est l'énergie. Ils obtiennent des ordres de grandeur à l'aide de simulations.

Nous nous intéressons ici d'abord à la distribution limite des énergies de structures secondaires non optimales, dont on montre qu'elle est gaussienne à l'aide du théorème des quasi-puissances, avec des paramètres (espérance et variance) calculables numériquement à l'aide de Maple. Nous ébauchons ensuite, suite à une idée de Jean-Marc Steyaert, grâce d'une part à cette étude, et d'autre part également à des simulations permettant de connaître les ordres de grandeur des énergies optimales, comment améliorer en complexité l'algorithme AMSAG, qui recherche la structure secondaire optimale d'un mot d'ARN donné.

5.1 Rappels

On peut représenter des structures secondaires par des arbres de dérivation de grammaire. On fait correspondre à chaque règle un attribut, correspondant physiquement à de l'énergie. On a indiqué des exemples d'attribut dans la figure 5.2 ci-dessous.

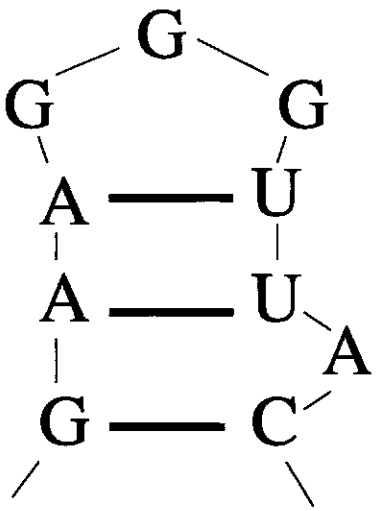


FIG. 5.1 – Portion de structure secondaire.

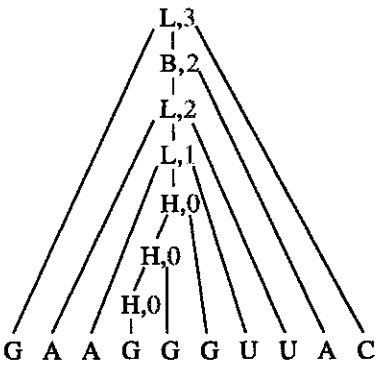


FIG. 5.2 – Arbre de dérivation correspondant.

5.2 Grammaire

Dans un mail daté du 27 décembre 2001, inspiré par [135], Jérôme Waldispühl propose la grammaire suivante :

```
Struct  -> Loop | * | Struct Loop | Struct *
Loop    -> ( Loop ) | ( Hairpin ) | ( LoopSeq ) | ( Internal )
Internal -> * Rgauche | Rdroite *  | * B_int *
Hairpin  -> Hairpin * | * * *
LoopSeq  -> Loop | * | LoopSeq Loop | LoopSeq *
Rgauche  -> * Rgauche | Loop
Rdroite  -> Rdroite * | Loop
B_int    -> * B_int | * B_int | Loop
```

Afin de disposer de notations plus légères, notons les éléments de l'alphabet de la grammaire ci-dessus par des lettres, en général l'initiale, sauf pour `LoopSeq` désigné par T comme Trèfle (même si on n'a pas forcément un trèfle), et pour `Rgauche` et `Rdroite` respectivement notés G

et D comme Gauche et Droite, ce qui se réécrit :

$$\begin{aligned}
S &\longrightarrow L \mid * \mid SL \mid S* \\
L &\longrightarrow (L) \mid (H) \mid (T) \mid (I) \\
I &\longrightarrow *G \mid D* \mid *B* \\
H &\longrightarrow H* \mid *** \\
T &\longrightarrow L \mid * \mid TL \mid T* \\
G &\longrightarrow *G \mid L \\
D &\longrightarrow D* \mid L \\
B &\longrightarrow B* \mid *B \mid L
\end{aligned}$$

Les parenthèses autour de M ou de S correspondent à une paire de bases appariées $A-U$, $U-A$, $G-C$, $C-G$, et chaque paire de parenthèses donnent en fait lieu à quatre dérivations. Les autres lettres correspondent respectivement à :

- S : boucle extérieure
- L : hélice ou échelle
- I : boucle interne de tout type
- H : épingle (hairpin)
- T : suite de boucles
- G : renflement à gauche (*bulge*), ou boucle proéminente
- D : renflement à droite
- B : boucle interne bilatérale, ou double renflement (à droite et à gauche)

Les étoiles $*$ se réécrivent chacune en une des lettres A, U, G ou C.

5.3 Énergie

La structure secondaire d'un ARN correspond à l'énergie libre maximale de la molécule. Cette énergie est négative, donc il s'agit de minimiser l'opposé de cette énergie. Afin de manipuler des grandeurs positives, on travaillera avec l'opposé de l'énergie libre, et lorsqu'on parlera d'énergie, il s'agira toujours de l'opposé de l'énergie libre. On note a et b les énergies $A-U$ et $G-C$. On reprend les notations de [212] pour les énergies (positives) des différentes sous-structures ($k \geqslant 1$) :

- $\xi(k) = \tau k + \tau_0 =$ énergie de fin de boucle *(hairpin)* de k bases
- $\beta(k) = \lambda k + \lambda_0 =$ énergie d'un *bulge* de k bases
- $\gamma(k) = \psi k + \psi_0 =$ énergie d'une boucle intérieure de k bases
- $\rho(k) = \phi k + \phi_0 =$ énergie de k bases non appariées dans une boucle multibranchée

De manière heuristique, en s'appuyant sur les tables de valeur données en *kcal/mole* dans [135] p.57-58, on trouve en interpolant avec Maple :

```
> #boucle proeminente
> fit[leastsquare[[x,y]]]([[1,2,3,4,5,6,7,8,9,10],
> [3.9,3.1,3.5,4.2,4.8,5.0,5.1,5.3,5.4,5.5]]);
                          y = 3.166666667 + .2569696970 x
```

```
> #epingle a cheveux
> fit[leastsquare[[x,y]]]([[3,4,5,6,7,8,9,10],
> [4.5,5.5,4.9,5.1,5.2,5.5,5.8,5.9]]);
                        y = 4.278571429 + .1571428571 x

> #boucle interne
> fit[leastsquare[[x,y]]]([[2,3,4,5,6,7,8,9,10],
> [4.1,4.5,4.9,5.3,5.7,5.8,6.0,6.1,6.2]]);
                        y = 3.810000000 + .2650000000 x
```

donc on prend $\tau = 4,279$, $\tau_0 = 0,157$, $\lambda = 3,167$, $\lambda_0 = 0,257$, $\psi = 3,810$, $\psi_0 = 0,265$.

En jetant un coup d'œil à la table correspondante, on prend une moyenne de $0,1$ pour l'énergie de mauvais appariement terminal, d'où $\phi = 0,1$ et $\phi_0 = 0$. En regardant la table des énergies des empilements, on prend en moyenne $a = e(AU) = -1$ et $b = e(GC) = -2$ pour les énergies de liaison. Il faudra raffiner ces valeurs plus tard, d'une part numériquement, et d'autre part qualitativement, en considérant effectivement des énergies d'empilement et pas seulement de liaison.

5.4 Série génératrice (SG)

5.4.1 Nature de la loi limite

On énumère les ARN par taille (nombre d'acides aminés) et opposé des énergies à l'aide des variables z et u. On note les SG avec les mêmes lettres et on trouve :

$$\begin{aligned}
S(z,u) &= L(z,u) + 4zu^{\phi+\phi_0} + S(z,u)L(z,u) + 4zu^{\phi}S(z,u)\\
L(z,u) &= 2z^2(u^a+u^b)(L(z,u)+H(z,u)+T(z,u)+I(z,u))\\
I(z,u) &= 4zu^{\lambda+\lambda_0}(G(z,u)+D(z,u)) + 4^2z^2u^{2\psi+\psi_0}B(z,u)\\
H(z,u) &= 4zu^{\tau}H(z,u) + 4^3z^3u^{3\tau+\tau_0}\\
T(z,u) &= L(z,u) + 4zu^{\phi+\phi_0} + T(z,u)L(z,u) + 4zu^{\phi}T(z,u)\\
G(z,u) &= 4zu^{\lambda}G(z,u) + L(z,u)\\
D(z,u) &= 4zu^{\lambda}D(z,u) + L(z,u)\\
B(z,u) &= 2\times 4zu^{\psi}B(z,u) + L(z,u)
\end{aligned}$$

On remarque $S = T$ et $G = D$. On résout du bas vers la haut.

S est une fonction de la forme $A(z,u) + B(z,u)\sqrt{C(z,u)}$, et admet en tant que fonction de z une singularité dominante $\rho(u)$ de type quadratique, racine de $z \mapsto C(z,u)$, et $z \mapsto A(z,u)$ et $z \mapsto B(z,u)$ sont analytiques sur un intervalle de la forme $[0, \rho(u)+\varepsilon]$. C'est une racine simple — sauf valeurs numériques exceptionnelles —, et on a :

$$[z^n]S(z) = f(\rho(u))\rho(u)^{-n}n^{-3/2}\left(1+O\left(\frac{1}{n}\right)\right).$$

D'après le théorème des quasi-puissances de Hwang[85], on a une limite gaussienne suivant une convergence en loi : si on note X_n l'énergie d'un ARN tiré uniformément parmi tous les ARN de taille n, et μ_n et σ_n^2 respectivement la moyenne et la variance de cette variable aléatoire, alors pour tous α, β, on a

$$\mathbf{Pr}\left[\alpha < \frac{X_n - \mu_n}{\sigma_n} < \beta\right] \xrightarrow{n\to\infty} \int_\alpha^\beta e^{-\frac{x^2}{2}}\,dx.$$

On trouve μ_n et σ_n par les formules

$$\mu_n = \frac{[z^n]S_u(z,1)}{[z^n]S(z,1)} \text{ et } \sigma_n^2 = \frac{[z^n](S_{u^2}(z,1) + S_u(z,1))}{[z^n]S(z,1)} - \left(\frac{[z^n]S_u(z,1)}{[z^n]S(z,1)}\right)^2.$$

5.4.2 Moyenne et écart-type

Énergie et nombre de liaisons $A-U$ et $G-C$

On peut affiner l'étude précédente prenant en compte le nombre de liaisons $A-U$ et $G-C$ présentes dans la structure secondaire et intervenant dans le calcul final de l'énergie de cette structure secondaire.

On calcule la moyenne de ces paramètres, et la matrice de covariance entre l'énergie de la molécule, le nombre de liaisons $A-U$, et le nombre de liaisons $G-C$, en introduisant d'autres variables v et w pour énumérer ces deux derniers paramètres, si bien qu'à la place de la série bivariée $L(z,u)$ on a notamment la série quadrivariée

$$L(z,u,v,w) = 2z^2((uv)^a + (uw)^b)(L(z,u,v,w) + H(z,u,v,w) + T(z,u,v,w) + I(z,u,v,w)).$$

Les grandeurs voulues se calculent en effectuant des dérivations par rapport à la variable idoine dans la série $S(z,u,v,w)$, en faisant $u = v = w = 1$, et en effectuant l'asymptotique de la série génératrice univariée obtenue.

En effectuant des calculs similaires aux précédents avec Maple, on trouve finalement :

```
a:=1000;b:=2000;
tau:=4279;tau0:=157;lam:=3167;lam0:=257;
psi:=3810;psi0:=265;phi:=100;phi0:=0;

B:=solve(B=2*z*u^(psi)*B+L,B);
DD:=solve(DD=z*u^(lam)*DD+L,DD);
G:=DD;
T:=solve(T=L+z*u^(phi+phi0)+z*T*L+z*u^phi*T,T);
H:=solve(H=z*u^(tau)*H+z^3*u^(3*tau+tau0),H);
II:=solve(II=z*u^(lam+lam0)+z^2*u^(2*phi+phi0)*B,II);

#changement par rapport à l'univarié
2*z^2*X*(L+H+T+II)*(z*u^100+z*L-1);
expr:=simplify(%);
LL:=solve(L*(z*u^100+z*L-1)=expr,L);
subs(X=1,LL[1]);
series(subs({X=1,u=1},LL[1]),z,5);
series(subs({X=1,u=1},LL[2]),z,5);

#On choisit la deuxième solution
L:=subs(X=(u*ua)^a+(u*ub)^b,LL[2]);

S:=solve(S=L+z*u^(phi+phi0)+z*S*L+z*u^phi*S,S);

S0:=subs({u=1,ua=1,ub=1},S);
```

```
SM:=proc(i,j,k)
 option remember;
 local res;
 #if i*j*k=0 then subs({u=1,ua=1,ub=1},S); fi;
 if i>0 then res:=diff(S,u$i); else res:=S;fi;
 if j>0 then res:=diff(res,ua$j);fi;
 if k>0 then res:=diff(res,ub$k);fi;
 subs({u=1,ua=1,ub=1},res);
end:

#Ajout de la librairit algolib
libname:="$HOME/AlgolibMaple7/algolib",libname:
libname:="/auto/algo/mnguyen/AlgolibMaple7/algolib",libname:
readlib(equivalent):

Mom:=proc(i,j,k)
 #option remember;
 local res;
 #if i*j*k=0 then subs({u=1,ua=1,ub=1},S); fi;
 if i=2 then res:=evalf(equivalent(SM(2,0,0)+SM(1,0,0),z,n,3));fi;
 if j=2 then res:=evalf(equivalent(SM(0,2,0)+SM(0,1,0),z,n,3));fi;
 if k=2 then res:=evalf(equivalent(SM(0,0,2)+SM(0,0,1),z,n,3));fi;
 if i<2 and j<2 and k<2 then res:=evalf(equivalent(SM(i,j,k),z,n,3));fi;
 res;
end:

num:=evalf(equivalent(SM(0,0,0),z,n,2));

Cov:=proc(i,j,k)
 #option remember;
 local res;
 if k=0 then res:=Mom(i,j,k)/num-Mom(1,0,0)*Mom(0,1,0)/num^2;
 elif j=0 then res:=Mom(i,j,k)/num-Mom(1,0,0)*Mom(0,0,1)/num^2;
 elif i=0 then res:=Mom(i,j,k)/num-Mom(0,1,0)*Mom(0,0,1)/num^2;fi;
 res:=evalf(gdev(res,n=infinity,4));
end:

Var:=proc(i,j,k)
 local res;
 #option remember;
 if k=2 then res:=Mom(i,j,k)/num-Mom(0,0,1)^2/num^2;fi;
 if j=2 then res:=Mom(i,j,k)/num-Mom(0,1,0)^2/num^2;fi;
 if i=2 then res:=Mom(i,j,k)/num-Mom(1,0,0)^2/num^2;fi;
 res:=evalf(gdev(res,n=infinity,4));
end:

Moy:=proc(i,j,k)
 local res;
```

```
 res:=Mom(i,j,k)/num;
 res:=evalf(gdev(res,n=infinity,4));
end:


#RESULTATS
X,Y,Z : energie, #liaisons AU, #liaisons GC

Moy(1,0,0);
                                    345.6109241   11142.49816      1
      714.0634242 n + 870.973221 - ------------ + ------------ + O(----)
                                       1/2            n           3/2
                                      n                         n

Moy(0,1,0);
                                    84.02983606   2359.967909      1
      173.6132405 n - 22.0308706 - ------------ + ------------ + O(----)
                                       1/2            n           3/2
                                      n                         n
Moy(0,0,1);
                                    168.0596706   4719.935723      1
       347.2264777 n - 44.061760 - ------------ + ------------ + O(----)
                                       1/2            n           3/2
                                      n                         n

Cov(1,1,0);
                  2                  3/2                      1/2
            .0020 n  + .01543978883 n    - 95452.1043 n + O(n   )

Cov(1,0,1);
                 2                  3/2                      1/2
           .0040 n  + .03087957766 n    + 156322.2593 n + O(n   )
Cov(0,1,1);
                 2                  3/2                       1/2
          .00096 n  + .007598892935 n    - 144351.6321 n + O(n   )


Var(2,0,0);
                  2                  3/2                      1/2
            .0084 n  + .06340477544 n    + 970613.845 n + O(n   )

Var(0,2,0);
                 2                  3/2                       1/2
          .00048 n  + .003775246171 n    + 101437.4229 n + O(n   )

Var(0,0,2);
                 2                  3/2                      1/2
           .0020 n  + .01500418350 n    + 405749.6933 n + O(n   )
```

Donc si on note $\mathbf{m}$ le vecteur de la limite des moyennes de l'énergie des ARN, du nombre de liaisons $A-U$, et du nombre de liaisons $G-C$, divisés par n et Σ la matrice de covariance de ces trois paramètres normalisés par $\sqrt{n}$, on a

$$\mathbf{m} = \begin{bmatrix} 714.063 \\ 173.613 \\ 347.226 \end{bmatrix}$$

et

$$\Sigma = \begin{bmatrix} 970.613 & -95.452 & 156.322 \\ -95.452 & 101.437 & -144.352 \\ 156.322 & -144.352 & 405.750 \end{bmatrix}.$$

Soit $\mathbf{x} = (X_1, X_2, X_3)$ la distribution normale trivariée $\mathcal{N}(\mathbf{m}, \Sigma)$ de paramètres $\mathbf{m}$ et Σ. Alors elle admet pour fonction densité la fonction

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}\,|\Sigma|^{1/2}} e^{-\frac{1}{2}{}^t(\mathbf{x}-\mathbf{m})\Sigma^{-1}(\mathbf{x}-\mathbf{m})},$$

et les fonctions cumulatives pour une variable donnée X_k, les valeurs des deux autres variables étant connues, ont pour expression

$$\begin{aligned} F_{X_k|X_i,\ \forall i\neq k}(x_k \mid x_i,\ \forall i \neq k) &= \mathbf{Pr}\left[X_k \neq x_k \;\middle|\; X_i = x_i\ \forall i \neq k\right] \\ &= \frac{\int_{-\infty}^{x_K} f_{X_1,X_2,X_3}(x_1,\cdots,x_{k-1},u_k,x_{k-1},\cdots,x_3)du_k}{f_{X_1,\cdots,X_{k-1},X_{k-1},\cdots,X_3}(x_1,\cdots,x_{k-1},x_{k-1},\cdots,x_3)}. \end{aligned}$$

5.4.3 Amélioration de la complexité de l'algo AMSAG ?

Premiers résultats

Grâce à l'algorithme AMSAG [207] (une version améliorée des algorithmes d'Earley et de Cocke-Kasami-Younger [2, 133, 144]), minimisant l'énergie de la molécule en cherchant les règles de dérivation optimales par le biais d'une programmation dynamique, on peut prévoir la structure secondaire de manière fiable, ainsi que trouver des structures secondaires sous-optimales (d'énergie proche de l'énergie minimale possible) pouvant également posséder une signification biologique. Cela généralise la recherche d'un plus grand sous-mot bien parenthésé d'un mot donné, puisqu'il s'agit de rechercher *grosso modo* le plus de liaisons possibles entre des acides aminés [177].

Afin d'améliorer la complexité de l'algo AMSAG, nous allons éviter à l'algorithme de considérer toutes les dérivations possibles menant à un mot ou sous-mot de la chaîne de départ. Pour ce faire, on utilise la distribution des énergies pour pouvoir reconnaître si une énergie est optimale ou non. Dès qu'un attribut a une valeur visiblement non optimale, on élimine la dérivation correspondante. Cela élimine non seulement la dérivation, mais aussi toutes les sous-dérivations qu'on aurait dû considérer en conservant la dérivation.

Nous allons écarter les énergies non optimales X_n des sous-structures secondaires de sous-mots de taille n avec la politique suivante dans l'algorithme AMSAG :

1. Calculer $\mu_n = \mu n$ et $\sigma_n^2 = \sigma^2 n$.
2. Calculer $X_n - \mu_n$.

3. Calculer $P = \dfrac{\sigma_n^2}{(X_n - \mu_n)^2}$.
4. Garder X_n si $P < f(n)$.

avec $f(n)$ une fonction à déterminer heuristiquement.

À ce jour, des simulations faites avec des fonctions de coupure $f(n)$ heuristiques, permettent de se débarrasser d'environ 50% des dérivations.

Néanmoins si l'algorithme semble globalement mieux marcher, si on engendre des mots d'ARN engendrés de manière complètement aléatoires, sur les mots d'ARN naturels il ne semble pas fonctionner plus vite. Cela signifie que les mots d'ARN ne sont pas distribués uniformément et que notre modèle est insuffisant.

Travaux futurs

Plusieurs pistes s'offrent à nous dans la poursuite de cette étude.

On pourrait affiner les résultats selon le mot d'ARN de départ. S'il semble encore irréalisable de connaître la distribution des énergies possibles d'un mot d'ARN donné, il devrait être possible de connaître, au moins asymptotiquement, la distribution des énergies possibles d'un mot d'ARN en fonction du nombre de lettres A, U, G et C, ou du nombre de liaisons $A-U$ ou $G-C$ intervenant dans le calcul de l'attribut de la structure secondaire. Pour cela, il faudrait étudier des distributions limites jointes, ce qui ne paraît pas trop difficile *a priori* puisqu'on ne fait apparaître les lettres A, U, G et C que dans une seule règle, au dernier moment, lorsqu'on remplace des parenthèses appariées par des liaisons. Cependant pour l'instant les valeurs numériques trouvées pour les matrices de covariance entre ces divers paramètres ne semblent pas engageants pour trouver de meilleurs résultats dans cette voie.

Il reste aussi à étendre cette étude à d'autres modèles. L'utilisation de packages de grammaires attribuées devraient permettre le calcul automatique de moyennes et de variances en introduisant seulement les règles de dérivation accompagnées de leur coût.

5.4.4 Figures

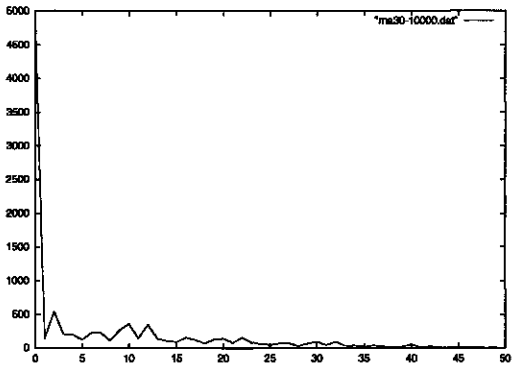


FIG. 5.3 – 10000 ARN de longueur 30.

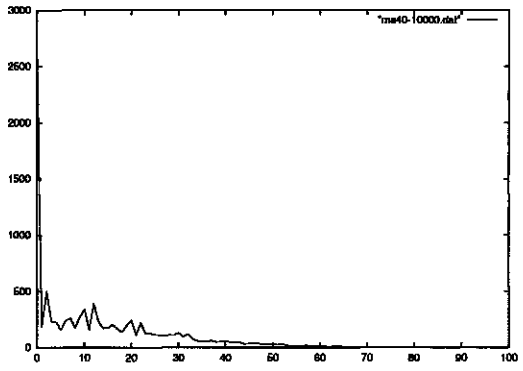


FIG. 5.4 – 10000 ARN de longueur 40.

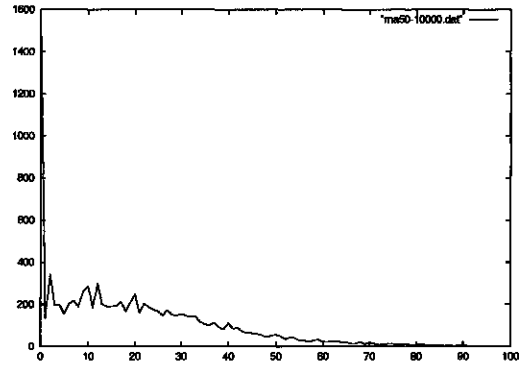


FIG. 5.5 – 10000 ARN de longueur 50.

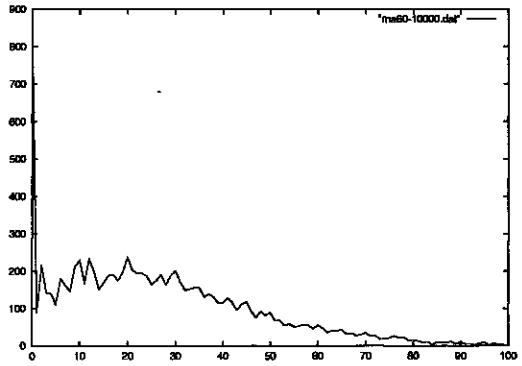


FIG. 5.6 – 10000 ARN de longueur 60.

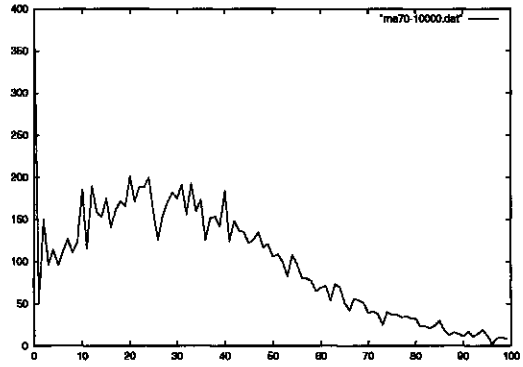


FIG. 5.7 – 10000 ARN de longueur 70.

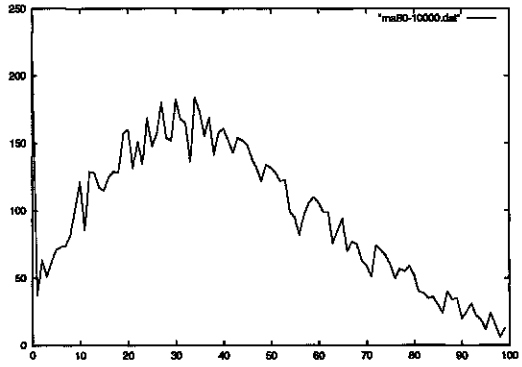


FIG. 5.8 – 10000 ARN de longueur 80.

5.5 Conclusion

Nous avons heuristiquement mis en évidence un algorithme de recherche de structures secondaires d'ARN qui fonctionne plus rapidement pour des mots d'ARN engendrés de manière aléatoire uniforme. Cependant il n'y a pas d'amélioration notable de l'algorithme lorsqu'on l'applique à des mots d'ARN issus du monde biologique, ce qui montre que dans notre problématique on ne peut pas considérer la distribution de ces mots comme uniforme.

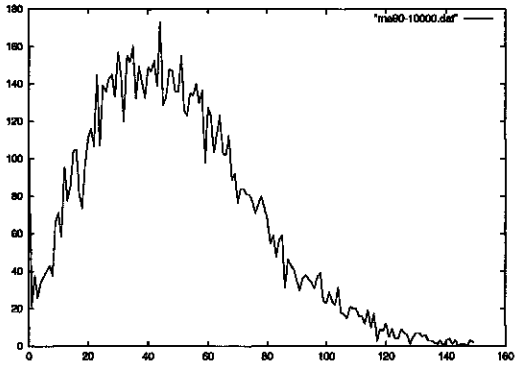


FIG. 5.9 – 10000 ARN de longueur 90.

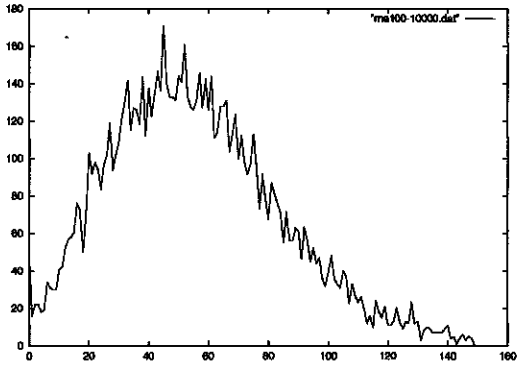


FIG. 5.10 – 10000 ARN de longueur 100.

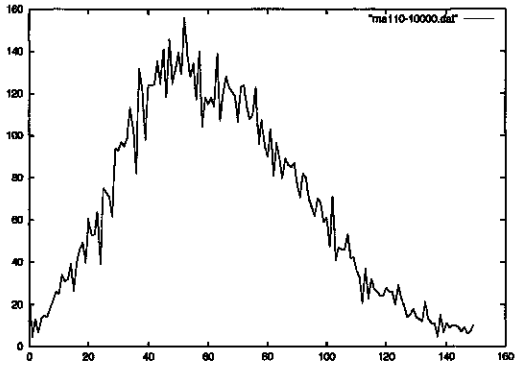


FIG. 5.11 – 10000 ARN de longueur 110.

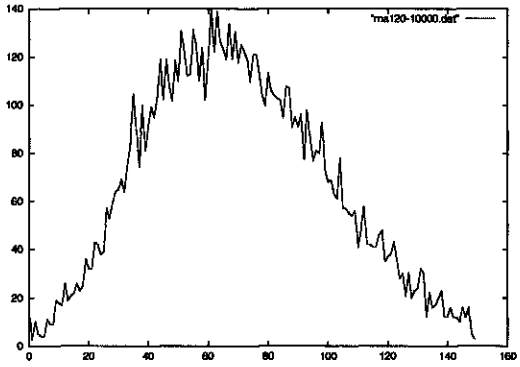


FIG. 5.12 – 10000 ARN de longueur 120.

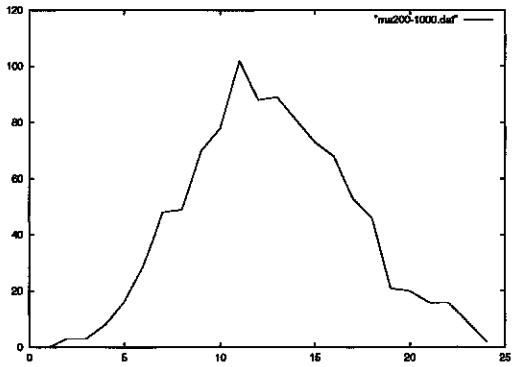


FIG. 5.13 – 1000 ARN de longueur 200.

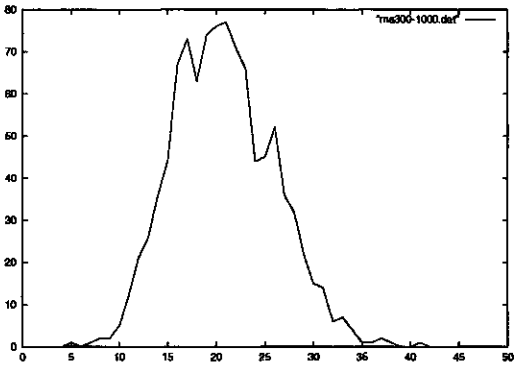


FIG. 5.14 – 1000 ARN de longueur 300.

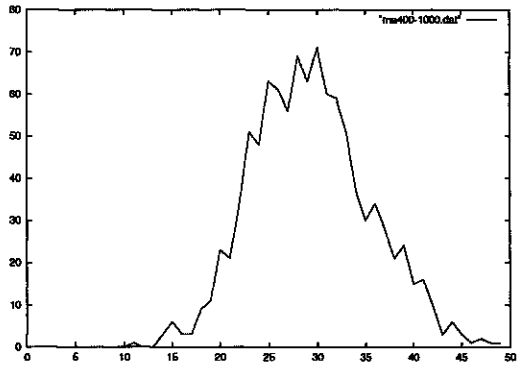


FIG. 5.15 – 1000 ARN de longueur 400.

Légende : Pour les graphes d'ARN de taille strictement inférieure à 200, une unité graphique représente une variation de $0,1$ $kcal/mol$, sinon à 1 $kcal/mol$. Le point d'abscisse x correspond au nombre d'ARN dont l'énergie minimale de repliement appartient à $[x, x+1[$.

Chapitre 6

Lois stables pour arbres avec opérateurs "−"

Le théorème central limite s'applique à des sommes de variables aléatoires identiques identiquement distribuées avec variance. Lorsqu'on somme des variables *i.i.d.* sans variance, on obtient des lois stables. Le lecteur non familiarisé avec les lois stables pourra consulter avec profit les ouvrages [22, 63, 186].

Dans ce chapitre, nous ébauchons une nouvelle preuve de ce fait en utilisant des méthodes d'analyse complexe dans des cas particuliers. Nous montrons tout d'abord que la technologie développée en appendice dans [6] permet d'obtenir des lois limites locales stables en sommant des lois de Zipf.

Nous montrons ensuite que cette technologie ne fonctionne pas si l'on veut sommer des variables pouvant prendre des valeurs positives ou négatives arbitrairement grandes en valeur absolue. Nous nous attachons donc à passer par des fonctions caractéristiques, pour montrer que nous obtenons encore des lois stables pour la distribution limite de la valuation d'arbres avec opérateurs "−" et des opérandes suivant une loi de Zipf sans variance.

Ce qui resterait à faire consisterait à généraliser le théorème des quasi-puissances pour traiter des lois sans variance, lorsqu'on a une singularité $z = \rho(u)$ sans variance.

6.1 Lois sans variance : préliminaires

Nous donnons un exemple de loi sans variance, à savoir des lois de Zipf, présentées dans [70]. Nous rappelons ensuite comment obtenir une expansion de la densité de lois sans variance.

6.1.1 Loi de Zipf

Dans [70, 87], on introduit la notion de polylogarithme généralisé :

Définition 6.1 *Le polylogarithme généralisé $Li_{\alpha,r}$ d'ordre fractionnaire r, où r est un entier et α un nombre complexe quelconque, est donné par :*

$$Li_{\alpha,r}(z) = \sum_{n\geqslant 1} (\log n)^r \frac{z^n}{n^\alpha}.$$

Théorème 6.1 *La fonction $Li_{\alpha,r}(z)$ est analytique dans le plan $\mathbb{C} \setminus [1, +\infty[$. Pour $\alpha \neq 1, 2, \cdots$, la fonction $Li_{\alpha,0}(z)$ admet le développement*

$$\begin{cases} Li_{\alpha,0}(z) \sim \Gamma(1-\alpha)t^{\alpha-1} + \sum_{j\geqslant 0} \frac{(-1)^j}{j!}\zeta(\alpha-j)t^j, \\ t = -\log z = \sum_{\ell=1}^{\infty} \frac{(1-z)^\ell}{\ell}. \end{cases} \tag{6.1}$$

quand $z \to 1$ dans le secteur (ε étant un réel arbitraire strictement positif)

$$-\pi + \varepsilon < \arg(1-z) < \pi - \varepsilon.$$

Pour $r > 0$, l'expansion singulière de $Li_{\alpha,r}(z)$ s'obtient par

$$Li_{\alpha,r}(z) = (-1)^r \frac{\partial^r}{\partial \alpha^r} Li_{\alpha,0}(z)$$

et correspond à une dérivation terme à terme de (6.1).

Définition 6.2 *La loi de Zipf généralisée de paramètre $s > 1$ est la loi de la variable aléatoire Z définie par :*

$$\mathbf{Pr}[Z = k] = \frac{1}{\zeta(s)} \frac{1}{k^s}.$$

Du théorème 6.1 se déduit la proposition suivante :

Proposition 6.1 *Soit g la série génératrice de probabilité de la loi de Zipf généralisée de paramètre $s > 1$, d'expression*

$$g(u) = \sum_{k\geqslant 1} \frac{1}{\zeta(s)} \frac{u^k}{k^s}.$$

Soit $\lambda = s - 1$. Astreignons u à rester dans le secteur

$$-\pi + \varepsilon < \arg(1-z) < \pi - \varepsilon,$$

où $\varepsilon > 0$. Alors, lorsque u est au voisinage de 1 :

(1) *Si $0 < \lambda < 1$, il existe $h_\lambda > 0$ tel que*

$$g(u) = 1 - h_\lambda (1-u)^\lambda + O(1-u).$$

(2) *Si $1 < \lambda < 2$, il existe $h_1, h_\lambda > 0$ tels que*

$$g(u) = 1 - h_1(1-u) + h_\lambda (1-u)^\lambda + O((1-u)^2).$$

(3) *Si $\lambda > 2$, alors il existe $h_1, h_2 > 0$ tels que*

$$g(u) = 1 - h_1(1-u) + h_2(1-u)^2 + O((1-u)^{\lambda'})$$

avec $\lambda' = \min(\lambda, 3)$.

6.1.2 Densités Stables

Dans [63] p.581-583 (Feller volume II), on trouve le lemme suivant :

Lemme 6.1 *Soit*

$$\Psi(\zeta) = -|\zeta|^{\lambda} \cdot e^{\pm i\pi\gamma/2}$$

le logarithme de la fonction caractéristique générique des lois stables, avec $\pm = sgn(\zeta)$ *et :*

$$|\gamma| \leq \begin{cases} \lambda & si \quad 0 < \lambda < 1 \\ 2 - \lambda & si \quad 1 < \lambda < 2 \end{cases}.$$

Notons $p(x;\lambda,\gamma)$ *la fonction densité correspondante. Alors :*

(1) *Pour* $x > 0$ *et* $0 < \lambda < 1$

$$p(x;\lambda,\gamma) = \frac{1}{\pi x}\sum_{k=1}^{\infty} \frac{\Gamma(k\lambda+1)}{k!}(-x^{-\lambda})^k \sin\frac{k\pi}{2}(\gamma-\lambda).$$

(2) *Pour* $x > 0$ *et* $1 < \lambda < 2$

$$p(x;\lambda,\gamma) = \frac{1}{\pi x}\sum_{k=1}^{\infty} \frac{\Gamma(k/\lambda+1)}{k!}(-x)^k \sin\frac{k\pi}{2\lambda}(\gamma-\lambda).$$

Les valeurs pour $x < 0$ *se déduisent de la formule*

$$p(-x;\lambda,\gamma) = p(x;\lambda,-\gamma).$$

Corollaire 6.1 *Si*

$$\Psi(\zeta) = -b|\zeta|^{\lambda} \cdot e^{\pm i\pi\gamma/2},$$

avec $b > 0$, *alors la densité* f *de la variable aléatoire correspondante s'écrit*

$$f(x) = b^{-1/\lambda}p(b^{-1/\lambda}x;\lambda,\gamma).$$

Démonstration. En utilisant le changement de variable $t = b^{1/\lambda}\zeta$, et en remarquant que t et ζ sont de même signe, on trouve

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\zeta x}e^{-b|\zeta|^{\lambda}\cdot e^{\pm i\pi\gamma/2}}\,d\zeta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ib^{-1/\lambda}tx}e^{-|t|^{\lambda}\cdot e^{\pm i\pi\gamma/2}}b^{-1/\lambda}\,dt \\ &= b^{-1/\lambda}p(b^{-1/\lambda}x;\lambda,\gamma). \end{aligned}$$

□

Les cas $\gamma = -\lambda$ pour $0 < \lambda < 1$ et $\gamma = \lambda - 2$ pour $1 < \lambda < 2$ se montreront d'un intérêt particulier pour la suite, si bien qu'on introduit la notation simplifiée $p(x;\lambda)$ définie par :

Définition 6.3 *Pour* $\lambda < 1$,

$$\begin{aligned} p(x;\lambda) &= p(x;\lambda,-\lambda) \\ &= \frac{1}{\pi x}\sum_{k=1}^{\infty} \frac{\Gamma(k\lambda+1)}{k!}(-1)^{k-1}x^{-k\lambda}\sin k\pi\lambda. \end{aligned}$$

Pour $1 < \lambda < 2$,

$$\begin{aligned} p(x;\lambda) &= p(x;\lambda,\lambda-2) \\ &= \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k/\lambda+1)}{k!}(-1)^{k-1}x^k\sin\frac{k\pi}{\lambda}. \end{aligned}$$

Remarque 6.1 *L'expression ci-dessus pour* $p(x;\lambda)$*, a priori valide seulement pour* $x > 0$*, est aussi valide pour* $x < 0$.

Démonstration. Soit $x < 0$. Alors $-x > 0$ et

$$\begin{aligned} p(x;\lambda) &= p(x;\lambda,\lambda-2) \\ &= p(-x;\lambda,2-\lambda) \\ &= \frac{1}{\pi(-x)}\sum_{k=1}^{\infty}\frac{\Gamma(k/\lambda+1)}{k!}(-(-x))^k\sin\frac{k\pi}{2\lambda}((2-\lambda)-\lambda) \\ &= -\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k/\lambda+1)}{k!}x^k\sin\frac{k\pi}{2\lambda}(2-2\lambda) \\ &= -\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k/\lambda+1)}{k!}x^k(-1)^k\sin\frac{k\pi}{\lambda} \\ &= \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k/\lambda+1)}{k!}(-1)^{k-1}x^k\sin\frac{k\pi}{\lambda}. \end{aligned}$$

□

6.2 Loi limite stable

6.2.1 Loi limite locale

Dans [6] appendice B, on trouve la phrase suivante :

"The sum of a large number of independent copies of the Zipf law of parameter $s \in (1,3)$ *satisfies a* local *limit law of the stable type with parameter* $s-1$. *"*

Nous nous proposons ici de détailler la preuve de ce fait. Notons que dans [6], on cherchait plutôt la taille n d'une structure, étant donné la taille k fixée d'une sous-structure. En outre, pour des exposants $\lambda < 1$, on semblait conduit à des lois stables d'index $1/\lambda$. Ici, de manière classique, on veut trouver, en sommant n variables *i.i.d.*, trouver la distribution limite de la somme k de ces variables, et on retrouve bien dans tous les cas λ comme index de la loi limite stable.

Proposition 6.2 *La somme de* n *variables i.i.d. suivant une loi de Zipf généralisée de paramètre* $s \neq 2$ *possède une loi limite locale stable de paramètre* $\min(2, s-1)$.

Démonstration. Notons $H(u)$ la série génératrice de probabilité de la loi de Zipf de paramètre $s = \lambda + 1$. Nous allons distinguer trois cas.

• Cas $0 < \lambda < 1$:

Supposons $H(u) = 1 - h_\lambda(1-u)^\lambda + O(1-u)$. Pour $k = xn^{1/\lambda}$, $x = O(1)$ (*i.e.* x restant dans un compact de $\mathbb{R}_+$), on veut calculer $[u^k]H^n(u) = \int H^n(u)u^{-k-1}\,du$.

Si on effectue le changement de variable $1-u = t/n^{1/\lambda}$, de telle sorte que t parcourt un contour de Hankel partant de $-\infty$, le contour initial pour u étant choisi à cet effet, on obtient d'une part

$$\begin{aligned} H^n &= \big(1-h_\lambda(1-u)^\lambda + O(1-u)\big)^n \\ &= \Big(1-h_\lambda\Big(\frac{t}{n^{1/\lambda}}\Big)^\lambda + O\Big(\frac{t}{n^{1/\lambda}}\Big)\Big)^n \\ &= \exp n\Big(-h_\lambda\frac{t^\lambda}{n} + O\Big(\max\Big(\frac{t^{2\lambda}}{n^{2\lambda}}, \frac{t}{n}\Big)\Big)\Big) \\ &= \exp\Big(-h_\lambda t^\lambda + O\Big(\max\Big(\frac{t^{2\lambda}}{n^{2\lambda-1}}, t\Big)\Big)\Big) \\ &\sim e^{-h_\lambda t^\lambda}, \end{aligned}$$

et d'autre part

$$\begin{aligned} u^{k+1} &= \Big(1-\frac{t}{n^{1/\lambda}}\Big)^{xn^{1/\lambda}+1} = e^{(xn^{1/\lambda}+1)\log\left(1-\frac{t}{n^{1/\lambda}}\right)} \\ &= e^{(xn^{1/\lambda}+1)\left(-\frac{t}{n^{1/\lambda}}+O\left(\frac{t^2}{n^{2/\lambda}}\right)\right)} \\ &\sim e^{-xt}. \end{aligned}$$

Donc

$$\begin{aligned} [u^k]H^n(u) &\sim \frac{1}{2i\pi n^{1/\lambda}}\int_{-\infty}^{0+} e^{-h_\lambda t^\lambda}e^{xt}\,dt \\ &\sim \frac{1}{2i\pi x n^{1/\lambda}}\int_{-\infty}^{0+} e^{-h_\lambda\frac{t^\lambda}{x^\lambda}+t}\,dt \\ &\sim \frac{1}{2i\pi x n^{1/\lambda}}\int_{-\infty}^{0+} e^t\sum_{j\ge 0}\frac{1}{j!}\Big(-h_\lambda\frac{t^\lambda}{x^\lambda}\Big)^j\,dt \\ &\sim \frac{1}{xn^{1/\lambda}}\sum_{j\ge 0}\frac{1}{j!\Gamma(-\lambda j)}\Big(-\frac{h_\lambda}{x^\lambda}\Big)^j \\ &\sim -\frac{1}{\pi x n^{1/\lambda}}\sum_{j\ge 0}\frac{\Gamma(1+\lambda j)}{j!}\sin(\pi\lambda j)\Big(-\frac{h_\lambda}{x^\lambda}\Big)^j \\ &\sim p(x/h_\lambda^{1/\lambda};\lambda)/(h_\lambda n)^{1/\lambda}, \end{aligned}$$

ce qui correspond à une loi stable de fonction caractéristique

$$t \mapsto e^{-h_\lambda|t|^\lambda e^{\mp i\pi\lambda}}.$$

• Cas $1<\lambda<2$:

Supposons $H(u) = 1-h_1(1-u)+h_\lambda(1-u)^\lambda+O((1-u)^2)$. Pour $k = h_1 n - xn^{1/\lambda}$, $x = O(1)$ (x dans un compact de $\mathbb{R}$), on veut calculer $[u^k]H^n(u) = \int H^n(u)u^{-k-1}\,du$. Si on effectue le

changement de variable $1-u=\dfrac{t}{n^{1/\lambda}}$, on obtient d'une part

$$\begin{aligned} H^n &= \big(1-h_1(1-u)+h_\lambda(1-u)^\lambda+O((1-u)^2)\big)^n \\ &= \Big(1-h_1\frac{t}{n^{1/\lambda}}+h_\lambda\Big(\frac{t}{n^{1/\lambda}}\Big)^\lambda+O\Big(\Big(\frac{t}{n^{1/\lambda}}\Big)^2\Big)\Big)^n \\ &= \Big(1-h_1\frac{t}{n^{1/\lambda}}+h_\lambda\frac{t^\lambda}{n}+O\Big(\frac{t^2}{n^{2/\lambda}}\Big)\Big)^n \\ &= \exp n\Big(-h_1\frac{t}{n^{1/\lambda}}+h_\lambda\frac{t^\lambda}{n}+O\Big(\frac{t^2}{n^{2/\lambda}}\Big)\Big)\sim\exp\Big(-h_1\frac{t}{n^{1/\lambda-1}}+h_\lambda t^\lambda\Big), \end{aligned}$$

et d'autre part,

$$\begin{aligned} u^{k+1} &= \Big(1-\frac{t}{n^{1/\lambda}}\Big)^{h_1n-xn^{1/\lambda}+1}=\Big(1-\frac{t}{n^{1/\lambda}}\Big)e^{h_1n-xn^{1/\lambda}\left(-\frac{t}{n^{1/\lambda}}+O\left(\frac{t^2}{n^{2/\lambda}}\right)\right)} \\ &\sim e^{-h_1\frac{t}{n^{1/\lambda-1}}+xt}. \end{aligned}$$

On effectue le changement de variable $w=h_\lambda t^\lambda$ de telle sorte que w parcourt un contour de Hankel partant de $-\infty$. Pour cela on a choisi pour t le parcours d'intégration suivant :

$$t-t_0=re^{\mp i\pi/\lambda},$$

avec $t_0=1$. Le parcours initial de u s'en déduit :

$$u-u_0=re^{\pm i\pi\gamma},$$

avec $u_0=1-t_0/n^{1/\lambda}$ et $\gamma=1-1/\lambda$. En effet :

$$\begin{aligned} u-u_0 &= re^{\pm i\gamma} \\ &= (1-tn^{-1/\lambda})-u_0 \\ &= (1-(t_0+re^{\mp i\pi\lambda})n^{-1/\lambda})-u_0 \\ &= 1-t_0n^{-1/\lambda}-u_0-re^{\mp i\pi/\lambda} \\ &= 1-t_0n^{-1/\lambda}-u_0+re^{\pm i\pi(1-/\lambda)}. \end{aligned}$$

Donc, si on pose $\alpha=1/\lambda$:

$$\begin{aligned} [u^k]H^n(u) &\sim \frac{1}{2i\pi n^{1/\lambda}}\int e^{h_\lambda t^\lambda-xt}\,dt \\ &\sim \frac{1}{2i\pi n^{1/\lambda}}\int_{-\infty}^{0+}e^{w-x\frac{\sigma^\alpha}{h_\lambda^\alpha}w^\alpha}\frac{\sigma^\alpha}{h_\lambda^\alpha}\alpha w^{\alpha-1}\,dw \\ &\sim \alpha\frac{\sigma^\alpha}{h_\lambda^\alpha}\frac{1}{2i\pi n^{1/\lambda}}\int_{-\infty}^{0+}e^{w-x\frac{\sigma^\alpha}{h_\lambda^\alpha}w^\alpha}\alpha w^{\alpha-1}\,dw \\ &\sim \frac{1}{2i\pi xn^{1/\lambda}}\int_{-\infty}^{0+}e^w\sum_{j\geq0}\frac{1}{j!}\Big(-x\frac{\sigma^\alpha}{h_\lambda^\alpha}w^\alpha\Big)^j\,dw \\ &\sim \frac{1}{xn^{1/\lambda}}\sum_{j\geq0}\frac{1}{j!\Gamma(-\alpha j)}\Big(-x\frac{\sigma^\alpha}{h_\lambda^\alpha}\Big)^j \\ &\sim -\frac{1}{xn^{1/\lambda}}\sum_{j\geq0}\frac{\Gamma(\alpha j+1)\sin(\pi\alpha j)}{j!}\Big(-x\frac{\sigma^\alpha}{h_\lambda^\alpha}\Big)^j \\ &\sim \frac{\sigma^{1/\lambda}}{h_\lambda^{1/\lambda}n^{1/\lambda}}p(-x\frac{\sigma^{1/\lambda}}{h_\lambda^{1/\lambda}};\lambda), \end{aligned}$$

ce qui correspond à une loi stable de fonction caractéristique

$$t \mapsto e^{-h_\lambda |t|^\lambda e^{\pm i\pi(\lambda-2)}} = e^{-h_\lambda |t|^\lambda e^{\pm i\pi\lambda}}.$$

• Cas $2 < \lambda$. Posons $\lambda' = \min(\lambda, 3)$.

Supposons $H(u) = 1 - h_1(1-u) + h_2(1-u)^2 + O((1-u)^{\lambda'})$. Pour $k = h_1 n + x\sqrt{n}$, $x = O(1)$ (x dans un compact de $\mathbb{R}_+$), on veut calculer $[u^k]H^n(u) = \int H^n(u) u^{-k-1}\, du$.

Si on effectue le changement de variable $1 - u = \dfrac{t}{n^{1/2}}$, on obtient

$$\begin{aligned} H^n &= \big(1 - h_1(1-u) + h_2(1-u)^2 + O((1-u)^{\lambda'}\big)^n \\ &= \Big(1 - h_1\frac{t}{\sqrt{n}} + h_2\frac{t^2}{n} + O\Big(\frac{t^{\lambda'}}{n^{\lambda'/2}}\Big)\Big)^n \\ &= \exp n\Big(-h_1\frac{t}{\sqrt{n}} + \Big(h_2 - h_1^2\Big)\frac{t^2}{n} + O\Big(\frac{t^{\lambda'}}{n^{\lambda'/2}}\Big)\Big) \\ &= \exp\Big(-h_1 t\sqrt{n} + \Big(h_2 - h_1^2\Big)t^2 + O\Big(\frac{t^{\lambda'}}{n^{\lambda'/2-1}}\Big)\Big) \sim \exp\Big(-h_1 t\sqrt{n} + \Big(h_2 - h_1^2\Big)t^2\Big) \end{aligned}$$

et

$$\begin{aligned} u^{k+1} &= \Big(1 - \frac{t}{\sqrt{n}}\Big)^{h_1 n + x\sqrt{n} + 1} = \Big(1 - \frac{t}{\sqrt{n}}\Big) e^{h_1 n + x\sqrt{n}\left(-\frac{t}{\sqrt{n}} + \frac{t^2}{n} + O\left(\frac{t^{\lambda'}}{n^{\lambda'/2}}\right)\right)} \\ &= \Big(1 - \frac{t}{\sqrt{n}}\Big) e^{h_1 n + x\sqrt{n}\left(-\frac{t}{\sqrt{n}} + \frac{t^2}{n} + O\left(\frac{t^{\lambda'}}{n^{\lambda'/2}}\right)\right)} \\ &= \Big(1 - \frac{t}{\sqrt{n}}\Big) e^{-h_1 t\sqrt{n} + h_1 t^2 - xt + O\left(\frac{t^{\lambda'}}{n^{\lambda'/2}}\right)} \\ &\sim \exp\Big(-h_1 t\sqrt{n} - xt\Big). \end{aligned}$$

Donc

$$\begin{aligned} [u^k]H^n(u) &\sim \frac{1}{2i\pi\sqrt{n}} \int_{-\infty}^{0+} \exp\Big(-h_1 t\sqrt{n} + \Big(h_2 - h_1^2\Big)t^2 + h_1 t\sqrt{n} + xt\Big)\, dt \\ &\sim \frac{1}{2i\pi\sqrt{n}} \int_{-\infty}^{0+} \exp\Big(tx + (h_2 + h_1 - h_1^2)t^2\Big)\, dt, \end{aligned}$$

ce qui mène bien à une loi limite normale de variance $h_2 + h_1 - h_1^2$. □

6.2.2 Limite faible

La preuve précédente n'est plus valide dans le cas où la série génératrice de probabilité n'est pas prolongeable dans un secteur au-delà de $u = 1$. On peut néanmoins trouver des limites faibles en faisant intervenir les séries génératrices des moments et les fonctions caractéristiques.

Loi de Zipf

On traite d'abord le cas précédent en sommant des lois de Zipf.

Proposition 6.3 *Soit une loi de Zipf de paramètre $s = \lambda + 1 \in]1,2[\cup]2,3[$, de série génératrice de probabilité F de la forme*

$$F(u) = 1 - f_\lambda (1-u)^\lambda + O((1-u))$$

si $s < 2$, et de la forme

$$F(u) = 1 - f_1(1-u) - f_\lambda (1-u)^\lambda + O((1-u)^2)$$

si $s > 2$. Alors si X_n est la somme de n variables i.i.d. suivant cette loi de Zipf, $X_n/n^{1/\lambda}$ si $\lambda < 1$, $(X_n - nf_1)/n^{1/\lambda}$ si $\lambda > 1$, converge en distribution vers une loi stable de fonction caractéristique

$$\phi(t) = \exp\left(-f_\lambda |t|^\lambda e^{\epsilon i \frac{\pi}{2}\lambda}\right),$$

où ϵ désigne le signe de t.

Démonstration. Traitons d'abord le cas $0 < \lambda < 1$. En effectuant le changement de variable $u = e^{it}$ on trouve

$$\begin{aligned} F(u) &= 1 - f_\lambda (1-u)^\lambda + O(1-u) \\ F(e^{it}) &= 1 - f_\lambda (1-e^{it})^\lambda + O(1-e^{it}) \\ &= 1 - f_\lambda (1-e^{it})^\lambda + O(t) \text{ car } e^{it} = \sum \frac{(it)^n}{n!}. \end{aligned}$$

Donc

$$\begin{aligned} F^n(e^{it}) &= \left(1 - f_\lambda (1-e^{it})^\lambda + O(t)\right)^n \\ &= \exp\left(n \log\left(1 - f_\lambda (1-e^{it})^\lambda + O(t)\right)\right) \\ &= \exp\left(n\left(-f_\lambda (1-e^{it})^\lambda + O(|t|^{2\lambda})\right)\right) \\ &= \exp\left(n\left(-f_\lambda (it + O((it)^2))^\lambda + O(|t|^{2\lambda})\right)\right) \\ &= \exp\left(n\left(-f_\lambda |t|^\lambda (\epsilon i + O(t))^\lambda + O(|t|^{2\lambda})\right)\right) \\ &= \exp\left(n\left(-f_\lambda |t|^\lambda e^{\epsilon i \frac{\pi}{2}\lambda}(1 + O(t))^\lambda + O(|t|^{2\lambda})\right)\right) \\ &= \exp\left(n\left(-f_\lambda |t|^\lambda e^{\epsilon i \frac{\pi}{2}\lambda} + O(|t|^{2\lambda})\right)\right), \end{aligned}$$

et

$$\begin{aligned} \phi_{X_n}(t) &= \exp\left(-nf_\lambda |t|^\lambda e^{\epsilon i \frac{\pi}{2}\lambda} + O(n|t|^{2\lambda})\right) \\ \phi_{X_n/n^{1/\lambda}}(t) &= \exp\left(-nf_\lambda \left(\frac{|t|}{n^{1/\lambda}}\right)^\lambda e^{\epsilon i \frac{\pi}{2}\lambda} + O\left(\frac{n|t||^{2\lambda}}{n^2}\right)\right) \\ &= \exp\left(-f_\lambda |t|^\lambda e^{\epsilon i \frac{\pi}{2}\lambda} + O\left(\frac{|t|^{2\lambda}}{n}\right)\right), \end{aligned}$$

d'où le résultat en passant à la limite.

Par des calculs similaires sur $F(e^{it})^n e^{-nf_1 t}$, on montre

$$\lim_{n\to\infty} \phi_{(X_n - nf_1)/n^{1/\lambda}}(t) = \exp\left(f_\lambda e^{\epsilon i \frac{\pi}{2}\lambda} |t|^\lambda\right).$$

□

Loi de Zipf symétrisée

On considère une variable aléatoire de départ de variance infinie pouvant prendre aussi bien des valeurs positives que des valeurs négatives, de telle sorte que sa série génératrice ne soit définie ni en-deça ni au-delà de $u = 1$. Nous allons montrer que les calculs précédents peuvent encore parfois marcher sur un exemple, alors qu'on ne peut plus montrer de loi limite locale stable en faisant appel à un contour de Hankel.

Proposition 6.4 *Soit F la série génératrice d'une loi de Zipf de paramètre $s = \lambda + 1 \in]1, 2[$. Soit X_n la somme de n variables i.i.d. de série génératrice de probabilité $(F(u) + F(1/u))/2$. Alors $X_n/n^{1/\lambda}$ converge en distribution vers une loi stable de fonction caractéristique*

$$\phi(t) = \exp\left(-\cos\left(\frac{\pi}{2}\lambda\right)|t|^\lambda\right),$$

où ϵ désigne le signe de t.

Démonstration. Supposons $0 < \lambda < 1$. Alors en utilisant le théorème 6.1, on voit que la série génératrice F admet un développement de la forme

$$F(u) = 1 - f_\lambda(1-u)^\lambda - f_1(1-u) - f_{1+\lambda}(1-u)^{1+\lambda} + O((1-u)^2),$$

donc

$$\begin{aligned} F(e^{it}) + F(e^{-it}) &= 2 - f_\lambda\left[(1-e^{it})^\lambda + (1-e^{-it})^\lambda\right] \\ &\quad + f_1\left[(1-e^{it}) + (1-e^{-it})\right] \\ &\quad + f_{1+\lambda}\left[(1-e^{it})^{1+\lambda} + (1-e^{-it})^{1+\lambda}\right] + O(t^2) \\ &= 2 - f_\lambda\left[(1-e^{it})^\lambda + (1-e^{-it})^\lambda\right] + O(t^2). \end{aligned}$$

Si $1 < \lambda < 2$, alors un calcul similaire donne le même résultat final.

Or pour tout t,

$$(1-e^{it})^\lambda + (1-e^{-it})^\lambda = 2\cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda + \lambda\sin\left(\frac{\pi\lambda}{2}\right)|t|^{1+\lambda} + O(t^2),$$

donc il existe $\lambda' > \lambda$ tel que

$$\frac{F(e^{it}) + F(e^{-it})}{2} = 1 - \cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda + O(|t|^{\lambda'})$$

et

$$\begin{aligned} \phi_{X_n}(t) &= \left(\frac{F(e^{it}) + F(e^{-it})}{2}\right)^n \\ &= \left(1 - \cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda + O(|t|^{\lambda'})\right)^n \\ &= \exp\left(n\log\left(1 - \cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda + O(|t|^{\lambda'})\right)\right) \\ &= \exp\left(n\cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda + O(|t|^{\lambda'})\right), \end{aligned}$$

d'où

$$\lim_{n\to\infty} \phi_{X_n/n^{1/\lambda}}(t) = \exp\left(\cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda\right).$$

□

6.3 Arbre –

Proposition 6.5 *Soit X_n la valuation des arbres de Catalan de taille n avec des opérateurs – aux nœuds internes et une loi de Zipf de paramètre $s = \lambda + 1 \in]1, 2[$ aux feuilles, de série génératrice de probabilité F de la forme*

$$F(u) = 1 - f_\lambda (1-u)^\lambda - f_1(1-u) - f_{1+\lambda}(1-u)^{1+\lambda} + O((1-u)^2).$$

Alors $X_n/n^{1/(1+\lambda)}$ converge en distribution vers une loi stable de fonction caractéristique

$$\phi(t) = \exp\left(\cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda\right).$$

Démonstration. Soit $T(z,u)$ la série génératrice énumérant les arbres par la taille selon z, et par valuation selon u. Rappelons que si F est la série génératrice des feuilles, et $\overline{F}(u) = F(1/u)$, alors on a

$$\begin{aligned} T(z,u) &= \frac{1 - z(F+\overline{F})}{2z} \\ &\quad - \frac{\sqrt{\left(1 - z(F+\overline{F}) + 2z\sqrt{F\overline{F}}\right)\left(1 - z(F+\overline{F}) - 2z\sqrt{F\overline{F}}\right)}}{2z} \\ &= \frac{1 - z(F+\overline{F})}{2z} \\ &\quad - \frac{\sqrt{\left(1 - z(\sqrt{F} - \sqrt{\overline{F}})^2\right)\left(1 - z(\sqrt{F} + \sqrt{\overline{F}})^2\right)}}{2z}. \end{aligned}$$

Soit $\lambda' = \min(1 + 2\lambda, 2)$. En utilisant les formules

$$\sqrt{1-u} = \sum_{n \geqslant 0} \binom{-\frac{1}{2}}{n} (-u)^n,$$

pour $u \to 1$ dans le secteur

$$-\pi + \varepsilon < \arg(1-z) < \pi - \varepsilon,$$

et

$$(1 - e^{it})^\alpha + (1 - e^{-it})^\alpha = 2\cos\left(\frac{\pi}{2}\alpha\right)|t|^\alpha + O(t^{\min(1+\alpha,2)})$$

pour tout $\alpha > 0$, on trouve $\lambda' > \lambda$ tel que ($\lambda' = \min(1, 2\lambda)$ si $\lambda < 1$ et $\lambda' = 2$ si $\lambda > 1$)

$$\frac{\sqrt{F}(e^{it}) + \sqrt{F}(e^{-it})}{2} = 1 - \frac{f_\lambda}{2}|t|^\lambda \cos\left(\frac{\pi}{2}\lambda\right) + O(|t|^{\lambda'})$$

et

$$\left(\frac{\sqrt{F}(e^{it}) + \sqrt{F}(e^{-it})}{2}\right)^2 = 1 - f_\lambda |t|^\lambda \cos\left(\frac{\pi}{2}\lambda\right) + O(|t|^{\lambda'}).$$

La singularité dominante de $T(z,u)$ est

$$\frac{1}{\left(\sqrt{F} + \sqrt{\overline{F}}\right)^2}.$$

En outre, on a $(1-e^{it})^\lambda = O\left(|t|^\lambda\right)$, donc, d'une part, pour t au voisinage de 0, $T(z,e^{it})$ est analytique en z, et d'autre part,

$$\left(\frac{\sqrt{F}(e^{it})-\sqrt{F}(e^{-it})}{\sqrt{F}(e^{it})+\sqrt{F}(e^{-it})}\right)^2 = O(|t|^{2\lambda}).$$

Par analyse de singularité, on obtient

$$[z^n]T(z,e^{it}) = \frac{4^n}{\sqrt{\pi n^3}}\left(1+O\left(|t|^\lambda\right)\right)\left(1-f_\lambda|t|^\lambda\cos\left(\frac{\pi}{2}\lambda\right)+O(|t|^{\lambda'})\right)^n\left(1+O\left(\frac{1}{n}\right)\right),$$

avec un terme d'erreur $O\left(\frac{1}{n}\right)$ uniforme en t.

Or

$$\phi_{X_n}(t) = \frac{[z^n]T(z,e^{it})}{[z^n]T(z,1)}.$$

On obtient donc pour $X_n/n^{1/\lambda}$, en suivant les mêmes calculs que précédemment, une loi limite stable de fonction caractéristique

$$\phi(t) = \exp\left(\cos\left(\frac{\pi\lambda}{2}\right)|t|^\lambda\right).$$

□

Notons qu'on obtient la même loi limite qu'une somme de variables aléatoires *i.i.d.* de série génératrice de probabilité $(F+\overline{F})/2$, car la moyenne $f_1 = F'(1)$, lorsqu'elle existe, n'intervient pas dans la normalisation. Lorsqu'on développe des séries telles que $\sqrt{F}$, on a toujours un même plus grand moment fractionnaire d'ordre $\lambda < 2$.

Si on manipulait des séries F à variance, alors on n'obtiendrait pas la même loi limite qu'une somme de variables aléatoires *i.i.d.* de série génératrice de probabilité $(F+\overline{F})/2$. En effet $\sqrt{F}$ ou $\overline{F}$ n'ont plus nécessairement la même variance que F, ce qui induit des changements. On a notamment vu dans le chapitre 1

$$\mathbf{Var}\left(\frac{\sqrt{F}+\sqrt{\overline{F}}}{2}\right) = \mathbf{Var}(F)+\frac{F'(1)^2}{2},$$

ce qui est différent de

$$\mathbf{Var}\left(\frac{F+\overline{F}}{2}\right) = F''(1)+F'(1) = \mathbf{Var}(F)+F'(1)^2.$$

6.4 Conclusion

Nous avons montré qu'on pouvait exhiber des lois limites stables dans le cadre de l'analyse de singularité, en ébauchant une généralisation du théorème des quasi-puissances. Il aurait été possible de poursuivre l'étude de l'apparition de lois stables avec les martingales ou les théorèmes de point fixe.

Deuxième partie

Longueur de cheminement et moment d'inertie

Chapitre 7

Mouvement brownien

Soit $(x(t))_{t\in\mathbb{R}_+}$ une trajectoire aléatoire dans $\mathbb{R}_+$ dépendant du temps t, typiquement un mouvement brownien $(W(t))_{t\in\mathbb{R}_+}$. On appelle fonctionnelle de x une variable aléatoire dépendant de la trajectoire x. Les fonctionnelles $\int_0^1 |W(t)|\,dt$ et $\int_0^1 W^2(t)\,dt$ — qu'on pourra appeler aire et fonctionnelle quadratique respectivement — pour le mouvement brownien $(W(t))_{t\in\mathbb{R}_+}$ ont abondamment été étudiées : mentionnons [14, 118, 101, 116]. Par commodité de langage, même si ce n'est pas l'usage, il nous arrivera également de parler de moment d'inertie à la place de fonctionnelle quadratique, puisque cela revient à calculer le moment d'inertie physique du mouvement considéré par rapport à l'axe des abscisses.

Dans ce chapitre, nous étudierons en particulier le pont brownien et l'excursion brownienne. Un pont brownien est un mouvement brownien partant de 0 au temps $t = 0$ et retournant sur l'axe des abscisses en un temps t — c'est donc en fait une portion de mouvement brownien. Une excursion brownienne est un pont brownien conditionné à rester au-dessus de l'axe des abscisses. Un pont brownien normalisé est un pont brownien qui retourne sur l'axe des abscisses au temps $t = 1$. De même, une excursion brownienne normalisée est une excursion qui retourne sur l'axe des abscisses au temps $t = 1$.

Nous considérons les transformées de Laplace

$$\phi_{12B}(x,y) = \mathbf{E}\left[\exp\left(-x\left(\int_0^1 |b(t)|\,dt\right) - y\left(\int_0^1 b(t)^2\,dt\right)\right)\right]$$

et

$$\phi_{12E}(x,y) = \mathbf{E}\left[\exp\left(-x\left(\int_0^1 e(t)\,dt\right) - y\left(\int_0^1 e(t)^2\,dt\right)\right)\right],$$

où $(b(t))_{0\le t\le 1}$ et $(e(t))_{0\le t\le 1}$ désignent respectivement le pont brownien et l'excursion brownienne normalisés, et nous introduisons les notations $\phi_{1B}(x) = \phi_{12B}(x,0)$, $\phi_{2B}(x) = \phi_{12B}(0,x)$, $\phi_{1E}(x) = \phi_{12E}(x,0)$, $\phi_{2E}(x) = \phi_{12E}(0,x)$.

Nous montrerons comment déterminer la distribution de l'aire et du moment d'inertie du pont brownien et de l'excursion brownienne à l'aide de la formule de Feynman-Kac [111] et du théorème de Louchard [140], permettant de passer du pont à l'excursion. Ces calculs sont des rappels, à l'exception du calcul de ϕ_{2E}, dont l'expression était néanmoins connue par une autre approche (voir plus bas).

En utilisant la même méthodologie, nous montrons qu'on a concernant ϕ_{12B} et ϕ_{12E} des formules du type :

$$\int_0^\infty e^{-\alpha t}\phi_{12B}\left(\frac{a}{2}t^{3/2}, \frac{b^2}{8}t^2\right)\frac{dt}{\sqrt{2\pi t}} = -\frac{1}{b^{1/2}}\frac{D_{a^2b^{-3}-\frac{2\alpha}{b}-\frac{1}{2}}\left(2ab^{-3/2}\right)}{D'_{a^2b^{-3}-\frac{2\alpha}{b}-\frac{1}{2}}\left(2ab^{-3/2}\right)}$$

et

$$\int_0^\infty (e^{-\alpha t}-1)\phi_{12E}\left(\frac{a}{2}t^{3/2},\frac{b^2}{8}t^2\right)\frac{dt}{\sqrt{2\pi t^3}} = \sqrt{b}\frac{D'_{a^2b^{-3}-\frac{2\alpha}{b}-\frac{1}{2}}\left(2ab^{-3/2}\right)}{D_{a^2b^{-3}-\frac{2\alpha}{b}-\frac{1}{2}}\left(2ab^{-3/2}\right)} - \sqrt{b}\frac{D'_{a^2b^{-3}-\frac{1}{2}}\left(2ab^{-3/2}\right)}{D_{a^2b^{-3}-\frac{1}{2}}\left(2ab^{-3/2}\right)}.$$

Notons en particulier que les résultats ci-dessous pour la fonctionnelle quadratique $\int_0^1 x^2(t)\,dt$ peuvent également se retrouver (voir [213] par exemple) avec des outils plus sophistiqués, *via* la formule de Girsanov appliquée aux processus de Bessel de dimension δ, à l'aide de laquelle on trouve la transformée de Laplace

$$\phi\left(\frac{x}{2}\right) = \left(\frac{\sqrt{x}}{\sin\sqrt{x}}\right)^{\frac{\delta}{2}},$$

si on conditionne le processus à avoir longueur 1. Or un pont brownien est un processus de Bessel de dimension 1 conditionné à retourner à zéro, et une excursion brownienne est un processus de Bessel de dimension 3 conditionné à retourner à zéro. Donc on obtient la transformée de Laplace de la distribution de la fonctionnelle quadratique pour le pont et l'excursion en faisant respectivement $\delta = 1$ et $\delta = 3$.

À la fin du chapitre, nous donnons également une méthode pour calculer le moment joint d'ordre (1,1) de l'aire et du moment d'inertie du pont brownien et de l'excursion.

L'aire et le moment d'inertie pour le pont brownien font encore l'objet d'une étude récente par Leonid Tolmatz [202, 204, 205, 206].

7.1 Rappels

Dans ce paragraphe, nous effectuons deux rappels sur des résultats classiques sur le mouvement brownien : la propriété de scaling, portant sur plusieurs types de mouvement brownien, et la formule de Feynman-Kac, pour le mouvement brownien standard.

7.1.1 Propriété de scaling

La propriété de scaling s'exprime ainsi :

Proposition 7.1 (propriété de scaling) *Soit W un mouvement brownien de longueur L, et $\widetilde{W}$ un mouvement brownien normalisé, $\widetilde{W}(u) = \left[W(u) \mid L = 1\right]$. Alors*

$$W(u) \overset{\mathcal{D}}{=} \sqrt{\ell}\,\widetilde{W}(u/\ell), \text{ avec } L = \ell.$$

Cette propriété de scaling reste vraie si on remplace W par respectivement un pont brownien ou une excursion de longueur L, donc se terminant sur l'axe des abscisses au temps $t = L$, et si on remplace $\widetilde{W}$ par respectivement un pont brownien ou une excursion normalisé(e). La longueur L devient alors une variable aléatoire représentant la longueur du pont ou de l'excursion.

La propriété de scaling se montre très utile lorsqu'on étudie des fonctionnelles de type puissance. En effet, si on connaît une transformée de Laplace ϕ d'une fonctionnelle puissance $(x(t))_{0\le t\le 1} \mapsto \int_0^1 |x(t)|^\alpha\,dt$ appliquée à un mouvement brownien $W(s)$, alors la fonctionnelle du

mouvement brownien normalisé admet une transformée de Laplace ϕ_N liée à ϕ de la manière suivante :

$$\begin{aligned}\phi(t) &= \mathbf{E}\left[e^{-\int_0^t |W(s)|^\alpha\, ds}\right] \\ &= \mathbf{E}\left[e^{-t^{1+\alpha/2}\int_0^1 |\widetilde{W}(s)|\, ds}\right] \\ \phi(t) &= \phi_N(t^{1+\alpha/2}).\end{aligned}$$

Encore une fois ce calcul et donc la dernière égalité restent valables pour le pont et l'excursion.

7.1.2 Formule de Feynman-Kac

Ce qu'on appelle "formule de Feynman-Kac" recouvre en fait plusieurs formules exprimant la même idée. Nous donnons ici une formule portant sur des doubles transformées de Laplace de fonctionnelles de mouvement brownien. Les hypothèses du théorème ci-dessous pourraient aussi être affaiblies, mais elles se révèleront suffisantes pour l'usage que nous en ferons. On trouvera un exposé sur cette formule dans [111].

Théorème 7.1 *Soient V une fonction paire positive continue et f une fonction bornée et à support compact. Alors l'équation*

$$-\frac{1}{2}u''(x) + (\alpha + V(x))\, u(x) = f(x)$$

admet pour unique solution bornée C^2 la fonction u définie par

$$u(x) = \mathbf{E}_x\left[\int_0^\infty e^{-\alpha t - \int_0^t V(W(s))\, ds} f(W(t))\, dt\right].$$

La notation $\mathbf{E}_x$ signifie que l'espérance de la grandeur entre crochets est prise en intégrant sur l'ensemble des mouvements browniens $W(s)$ partant de x au temps $t = 0$.

7.2 Pont brownien

Le théorème ci-dessous est montré dans [190] pour le cas des fonctionnelles du type $\int_0^1 V(x(t))\, dt$ avec $V : x \mapsto |x|$ et $V : x \mapsto x^2$, mais en reprenant la même preuve, on peut généraliser aux cas des fonctionnelles $\int_0^1 V(x(t))\, dt$ associées à toute fonction V paire positive continue :

Théorème 7.2 (Shepp) *Soient V une fonction paire positive continue et g une solution bornée à l'infini de :*

$$-\frac{1}{2}u''(x) + (\alpha + V(x))\, u(x) = 0. \tag{7.1}$$

Alors la fonction

$$\phi(t) = \mathbf{E}_0\left[e^{-\int_0^t V(W(s))\, ds} \mid W(t) = 0\right]$$

vérifie

$$\int_0^\infty e^{-\alpha t}\phi(t)\, \frac{dt}{\sqrt{t}} = -\sqrt{2\pi}\frac{g(0)}{g'(0)}.$$

Démonstration. Soient g_1 et g_2 les deux solutions de (7.1) telles que g_1 est bornée en $+\infty$, g_2 est bornée en $-\infty$, et

$$g_1 g_2' - g_1' g_2 \equiv 2.$$

Alors pour toute fonction f bornée et à support compact, l'unique solution bornée u de

$$-\frac{1}{2}u''(x) + (\alpha + V(x))\, u(x) = f(x)$$

est donnée par l'opérateur de Green appliquée à f :

$$u(x) = g_1(x) \int_{-\infty}^{x} g_2(t) f(t)\, dt + g_2(x) \int_{x}^{\infty} g_1(t) f(t)\, dt,$$

et la formule de Feynman-Kac permet d'affirmer que

$$u(x) = \mathbf{E}_x \left[\int_0^\infty e^{-\alpha t - \int_0^t V(W(s))\, ds} f(W(s))\, dt \right],$$

où $W(s)$ est un mouvement brownien passant par x au temps $s = 0$.

En faisant $x = 0$, on obtient

$$\mathbf{E}_0 \left[\int_0^\infty e^{-\alpha t - \int_0^t V(W(s))\, ds} f(W(s))\, dt \right] = g_1(0) \int_{-\infty}^{0} g_2(t) f(t)\, dt + g_2(0) \int_0^\infty g_1(t) f(t)\, dt.$$

Pour obtenir un pont brownien à partir du mouvement brownien standard W, *i.e.* un mouvement brownien se terminant sur l'axe des abscisses au temps t, on choisit

$$f(x) = \frac{\sqrt{2\pi}}{2\varepsilon} \chi(|x| < \varepsilon),$$

où $\chi(|x| < \varepsilon)$ vaut 1 si $|x| < \varepsilon$, 0 sinon, et on fait $\varepsilon \to 0$.

Pour le membre de gauche, on obtient

$$\lim_{\varepsilon \to 0} \int_0^\infty \exp\left\{ -\alpha t - \int_0^t V(W(s))\, ds \right\} \frac{\chi(|W(t)| < \varepsilon)}{\mathbf{Pr}(|W(t)| < \varepsilon)} \frac{\mathbf{Pr}(|W(t)| < \varepsilon)}{\frac{2\varepsilon}{\sqrt{2\pi t}}} \frac{dt}{\sqrt{t}}.$$

Comme le rapport

$$\frac{\mathbf{Pr}(|W(t)| < \varepsilon)}{\frac{2\varepsilon}{\sqrt{2\pi t}}} \leq 1$$

tend vers 1 quand $\varepsilon \to 0$, le théorème de convergence dominée donne

$$\int_0^\infty e^{-\alpha t} \mathbf{E}_0 \left[e^{-\int_0^t V(W_s)\, ds} \middle|\, W(t) = 0 \right] \frac{dt}{\sqrt{t}} = \sqrt{2\pi} g_1(0) g_2(0).$$

Par symétrie, $g_1(x) \equiv g_2(-x)$, donc $g_1(0) = g_2(0)$ et $g_1'(0) = -g_2'(0)$. L'égalité $g_1 g_2' - g_1' g_2 \equiv 2$ écrite en zéro donne donc $-g_2(0) g_1'(0) = 1$, et

$$\int_0^\infty e^{-\alpha t} \mathbf{E}_0 \left[e^{-\int_0^t V(W(s))\, ds} \middle|\, W(t) = 0 \right] \frac{dt}{\sqrt{t}} = -\sqrt{2\pi} \frac{g_1(0)}{g_1'(0)}.$$

Si g est une solution bornée en $+\infty$ de (7.1), g est proportionnelle à g_1, et on a encore

$$\int_0^\infty e^{-\alpha t} \mathbf{E}_0 \left[e^{-\int_0^t V(W(s))\, ds} \middle|\, W(t) = 0 \right] \frac{dt}{\sqrt{t}} = -\sqrt{2\pi} \frac{g(0)}{g'(0)}.$$

□

7.2.1 Aire

En appliquant le théorème précédent à la fonction $V : x \mapsto |x|$, on obtient la double transformée de Laplace de la distribution de la fonctionnelle aire du pont brownien normalisé :

Proposition 7.2 (Shepp) *La transformée de Laplace ϕ_{1B} de la distribution de l'aire du pont brownien normalisé $(b(s))_{s\in[0,1]}$ vérifie*

$$\int_0^\infty e^{-ut}\phi_{1B}(\sqrt{2}s^{3/2})s^{-1/2}\,ds = -\sqrt{\pi}\frac{Ai(u)}{Ai'(u)}. \tag{7.2}$$

Démonstration. On s'intéresse à la fonction $V(x) = |x|$. Par propriété de scaling, on a

$$\mathbf{E}\left[e^{-\int_0^t |W(s)|\,ds} \mid W(t) = 0\right] = \mathbf{E}\left[e^{-t^{3/2}\int_0^1 |W(s)|\,ds} \mid W(1) = 0\right] = \mathbf{E}\, e^{-t^{3/2}\int_0^1 |b(s)|\,ds}.$$

Une solution bornée g à l'infini de

$$-\frac{1}{2}u''(x) + (\alpha + |x|)\,u(x) = 0,$$

admet, pour $x \geq 0$, l'expression

$$g(x) = Ai(2^{1/3}(x+\alpha)).$$

L'expression de g pour $x \leq 0$ est différente, mais nous n'avons pas besoin de la calculer : seules les valeurs en 0 de g et g' nous intéressent, et nous avons $g(0-) = g(0+)$ et $g'(0-) = g'(0+)$.

La transformée de Laplace $\phi_{1B}(u) = \mathbf{E}\left[e^{-u\int_0^1 |\widetilde{W}_s|\,ds}\right]$ vérifie :

$$\begin{aligned}\int_0^\infty e^{-\alpha t}\phi_{1B}(t^{3/2})\frac{dt}{\sqrt{t}} &= -\sqrt{2\pi}\frac{Ai(2^{1/3}(\alpha))}{Ai'(2^{1/3}(\alpha))}\\ \int_0^\infty e^{-us}\phi_{1B}(\sqrt{2}s^{3/2})s^{-1/2}\,ds &= -\sqrt{\pi}\frac{Ai(u)}{Ai'(u)}.\end{aligned}$$

□

7.2.2 Fonctionnelle quadratique

Dans ce paragraphe, nous montrons comment déterminer la distribution de la fonctionnelle quadratique du pont brownien à l'aide du théorème 7.2. Pour cette fonctionnelle, on peut trouver une expression sympathique pour la transformée de Laplace ϕ_{2B} : on n'a pas seulement à se contenter d'une double transformée de Laplace. Shepp avait montré comment obtenir cette double transformée de Laplace, mais a déterminé ϕ_{2B} en ayant recours à une autre méthode. Nous montrerons dans la preuve du lemme 7.1 comment procéder pour récupérer ϕ_{2B} à l'aide de la double transformée de Laplace. L'expression de ϕ_{2B} avait déjà été trouvé par Marc Kac [120, 119], en utilisant des formules assez compliquées, notamment *via* la théorie des expansions des fonctions propres, appliquée au fonctions paraboliques cylindriques.

Proposition 7.3 *La transformée de Laplace ϕ_{2B} de la distribution de la fonctionnelle quadratique du pont brownien normalisé vérifie*

$$\phi_{2B}\left(\frac{\lambda}{2}\right) = \left(\frac{\sqrt{\lambda}}{\sinh(\sqrt{\lambda})}\right)^{\frac{1}{2}}.$$

Démonstration. On s'intéresse à la fonction $V(x) = x^2$. Par propriété de scaling,

$$\mathbf{E}\left[e^{-\int_0^t W(s)^2\,ds} \mid W(t) = 0\right] = \mathbf{E}\left[e^{-t^2\int_0^1 W(s)^2\,ds} \mid W(1) = 0\right] = \mathbf{E}\left[e^{-t^2\int_0^1 b(s)^2\,ds}\right].$$

Soit D_ν la fonction parabolique cylindrique habituelle, dont le lecteur pourra trouver une définition dans [1]. C'est une solution bornée en $+\infty$ de

$$u'' = \left(\frac{1}{4}x^2 - \nu - \frac{1}{2}\right)u.$$

Si on pose

$$\xi_2 = \int_0^1 b(s)^2\,ds,$$

alors

$$\phi_{2B}(s) = \mathbf{E}\left[e^{-\xi_2 s}\right] = \mathbf{E}\left[e^{-s\int_0^1 W(t)^2\,dt}\right],$$

et d'après le théorème 7.2, on a

$$\int_0^\infty e^{-\alpha t}\phi_{2B}\left(\frac{1}{8}t^2\right)\frac{dt}{\sqrt{t}} = -\sqrt{2\pi}\frac{D_\nu(0)}{D'_\nu(0)} = \sqrt{\pi}\frac{\Gamma(\alpha+1/4)}{\Gamma(\alpha+3/4)}, \text{ avec } \nu = -\frac{1}{2} - 2\alpha.$$

Or en consultant une table de transformées de Laplace. comme dans [34], on trouve en particulier la formule, $\mathcal{L}$ désignant la transformation de Laplace :

$$\mathcal{L}\left(t \mapsto \left(1 - e^{-t/a}\right)^{\beta-1}\right) = p \mapsto aB(ap,\beta) = p \mapsto \frac{a\Gamma(ap)\Gamma(\beta)}{\Gamma(ap+\beta)}. \tag{7.3}$$

(Nous donnons une démonstration de cette formule dans le lemme en-dessous).

En utilisant $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $a = 1$, $\beta = \frac{1}{2}$ et $p = \alpha + \frac{1}{4}$, on trouve finalement

$$\phi_{2B}\left(\frac{\lambda}{2}\right) = \left(\frac{\sqrt{\lambda}}{\sinh(\sqrt{\lambda})}\right)^{\frac{1}{2}}. \tag{7.4}$$

□

Démontrons la formule (7.3), résumée dans le lemme en-dessous, de manière constructive, c'est-à-dire montrons que l'inverse de la transformation de Laplace de la fonction

$$p \mapsto aB(ap,\beta)$$

est donnée par la fonction

$$t \mapsto \left(1 - e^{-t/a}\right)^{\beta-1}.$$

Lemme 7.1 *La transformée de Laplace de la fonction* $t \mapsto \left(1 - e^{-t/a}\right)^{\beta-1}$ *est donnée par la formule*

$$\mathcal{L}\left(\left(1 - e^{-t/a}\right)^{\beta-1}\right) = aB(ap,\beta) = \frac{a\Gamma(ap)\Gamma(\beta)}{\Gamma(ap+\beta)}.$$

Démonstration. Il suffit d'appliquer la formule d'inversion de la transformation de Laplace et la formule des résidus. Les résidus de la fonction Γ sont donnés par

$$\mathrm{Res}(\Gamma, -n) = \frac{(-1)^n}{n!},$$

car pour tout n

$$\Gamma(-n+h) \underset{h\to 0}{\sim} \frac{(-1)^n}{n!h}.$$

Donc

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{ts}\frac{a\Gamma(as)\Gamma(\beta)}{\Gamma(as+\beta)}\,ds = \frac{1}{2\pi i}\int_{ac-i\infty}^{ac+i\infty} e^{tu/a}\frac{\Gamma(u)\Gamma(\beta)}{\Gamma(u+\beta)}\,du = \sum_{n\geq 0}(-e^{-t/a})^n\frac{\Gamma(\beta)}{n!\Gamma(-n+\beta)},$$

et pour conclure il suffit d'appliquer la formule du binôme de Newton

$$(1-e^{-t/a})^{\beta-1} = \sum_{n\geq 0}\binom{\beta-1}{n}(-e^{-t/a})^n = \sum_{n\geq 0}\frac{\Gamma(\beta)(-e^{-t/a})^n}{\Gamma(n+1)\Gamma(-n+\beta)}.$$

□

7.2.3 Aire et moment d'inertie joints

Trouver une expression de la double transformée de Laplace *via* la formule de Feynman-Kac requiert la connaissance d'une solution explicite de

$$u'' - 2(\alpha + V)u = 0,$$

où α est une constante. Or on sait résoudre ce type d'équation pour V polynôme du second degré [1]. Si on considère les coefficients de V comme des variables, cela revient à étudier la distribution jointe de l'aire et du moment d'inertie. On obtient donc un résultat du type

Proposition 7.4 *Soit ϕ_{12} la distribution jointe de l'aire et du moment d'inertie $|x|$ et x^2 du pont brownien normalisé. Si D_ν désigne la solution bornée en $+\infty$ de*

$$y'' = \left(\frac{1}{4}x^2 - \nu - \frac{1}{2}\right)y,$$

alors ϕ_{12} vérifie

$$\int_0^\infty e^{-\alpha t}\phi_{12}\left(\frac{a}{2}t^{3/2}, \frac{b^2}{8}t^2\right)\frac{dt}{\sqrt{t}} = -\frac{\sqrt{2\pi}}{b^{1/2}}\frac{D_\nu\left(ab^{-3/2}\right)}{D'_\nu\left(ab^{-3/2}\right)}. \tag{7.5}$$

Démonstration. Considérons l'équation différentielle

$$u'' - 2(\alpha + V)u = 0$$

avec $V(x) = \frac{b^2}{8}x^2 + \frac{a}{2}|x|$. Pour $x \geq 0$, elle s'écrit

$$u'' - \left(\frac{b^2}{4}x^2 + ax + 2\alpha\right)u = 0,$$

soit

$$u'' - \left(\frac{1}{4}\left(bx + \frac{2a}{b}\right)^2 - \frac{a^2}{b^2} + 2\alpha\right) u = 0.$$

Cherchons une solution de cette équation. D_ν est solution bornée en $+\infty$ de

$$y'' = \left(\frac{1}{4}x^2 - \nu - \frac{1}{2}\right) y$$

Donc si on pose $y(x) = D_\nu(px+q)$,

$$\begin{aligned} y''(x) &= p^2 D_\nu''(px+q) \\ &= p^2\left(\frac{1}{4}(px+q)^2 - \nu - \frac{1}{2}\right) y(x) \\ &= p^2\left(\frac{1}{4}p^2q^2 + \frac{1}{2}pqx + \frac{q^2}{4} - \nu - \frac{1}{2}\right) \\ &= \left(\frac{1}{4}p^4q^2 + \frac{1}{2}p^3qx + \frac{p^2q^2}{4} - p^2\left(\nu - \frac{1}{2}\right)\right) \end{aligned}$$

Par identification on trouve, pour respectivement p, q et ν,

- $p^4 = b^2$ d'où $p = b^{1/2}$,
- $p^3 q = a$ d'où $q = ap^{-3} = ab^{-3/2}$,
- $\dfrac{p^2q^2}{4} - p^2\left(\nu - \dfrac{1}{2}\right) = 2\alpha$ d'où

$$\begin{aligned} & \frac{p^2q^2}{4} &&= 2\alpha + p^2\left(\nu + \tfrac{1}{2}\right) \\ \iff & p^2\left(\nu + \tfrac{1}{2}\right) &&= \frac{p^2q^2}{4} - 2\alpha \\ \iff & \nu + \tfrac{1}{2} &&= \frac{q^2}{4} - \frac{2\alpha}{p^2} \\ \iff & \nu &&= \frac{q^2}{4} - \frac{2\alpha}{p^2} - \frac{1}{2} \\ \iff & \nu &&= \frac{a^2b^{-3}}{4} - \frac{2\alpha}{b} - \frac{1}{2}. \end{aligned}$$

Donc une solution est

$$\begin{aligned} \phi(x) &= c_0 D_\nu(px+q) \\ &= c_0 D_{\frac{a^2b^{-3}}{4} - \frac{2\alpha}{b} - \frac{1}{2}}\left(b^{1/2}x + ab^{-3/2}\right), \end{aligned}$$

où c_0 est une constante, de valeur indifférente par rapport à notre problème.

Si ϕ_{12} est définie par

$$\phi_{12}(a,b) = \mathbf{E}\left[e^{-\left(a\int_0^1 |\tilde{W}_s|\, ds + b\int_0^1 \tilde{W}_s^2\, ds\right)}\right],$$

il vient

$$\begin{aligned} \mathbf{E}\left[e^{-\int_0^t a|W(s)| + b^2 W(s)^2\, ds} \,\middle|\, W(t) = 0\right] &= \mathbf{E}\left[e^{-\left(t^{3/2}a\int_0^1 |\tilde{W}_s|\, ds + b^2t^2\int_0^1 \tilde{W}_s^2\, ds\right)}\right] \\ &= \phi_{12}\left(at^{3/2}, b^2t^2\right), \end{aligned}$$

et en utilisant le théorème 7.2, on obtient finalement

$$\int_0^\infty e^{-\alpha t}\phi_{12}\left(\frac{a}{2}t^{3/2}, \frac{b^2}{8}t^2\right)\frac{dt}{\sqrt{t}} = -\frac{\sqrt{2\pi}}{b^{1/2}}\frac{D_\nu\left(ab^{-3/2}\right)}{D_\nu'\left(ab^{-3/2}\right)}.$$

□

7.3 Excursion brownienne

Dans [140], on trouve le résultat suivant, permettant, connaissant la double transformée de Laplace de la distribution d'une fonctionnelle du pont brownien, d'en déduire la double transformée de Laplace de la distribution de la même fonctionnelle, appliquée à l'excursion brownienne.

Théorème 7.3 (Louchard) *On note $W(s)$ le mouvement brownien standard, $e(s)$ l'excursion brownienne, et L la durée de celle-ci. Soit $h \geq 0$ une fonction paire continue par morceaux et soit*

$$g(\alpha) = \int_0^\infty e^{-\alpha t} \mathbf{E}_0 \left[e^{-\int_0^t h(W(s))\,ds} \;\middle|\; W(t) = 0 \right] \frac{dt}{\sqrt{2\pi t}}.$$

Alors

$$\int_0^\infty \left(e^{-\alpha t} - 1\right) \mathbf{E}_0 \left[e^{-\int_0^t h(e_s)\,ds} \;\middle|\; L = t \right] \frac{dt}{\sqrt{2\pi t^3}} = -\frac{1}{g(\alpha)} + \lim_{\alpha \to 0} \frac{1}{g(\alpha)}.$$

Cette formule nous permet de connaître les lois de l'aire et du moment d'inertie de l'excursion brownienne.

7.3.1 Aire sous l'excursion

Proposition 7.5 (Louchard) *La transformée de Laplace ϕ_{1E} de la distribution de l'aire sous l'excursion est donnée par*

$$\phi_{1E}(z) = \sqrt{2\pi} z \sum_1^\infty e^{a_k (z/\sqrt{2})^{2/3}}, \tag{7.6}$$

où les a_k désignent les zéros consécutifs (par ordre décroissant sur la demi-droite négative de l'axe des abscisses) de la fonction d'Airy Ai.

Démonstration. La transformée de Laplace ϕ_B de la distribution de l'aire du pont brownien vérifie

$$\int_0^\infty e^{-\alpha t} \phi_B(t^{3/2}) \frac{dt}{\sqrt{t}} = -\sqrt{2\pi} \frac{Ai(2^{1/3}(\alpha))}{Ai'(2^{1/3}(\alpha))},$$

donc d'après le théorème 7.3 on a

$$\int_0^\infty \left(e^{-\alpha t} - 1\right) \mathbf{E} \left[e^{-t^{3/2} \int_0^1 e(s)\,ds} \right] \frac{dt}{\sqrt{2\pi t^3}} = \frac{2^{1/3} Ai'(2^{1/3} x)}{Ai(2^{1/3} x)} - \frac{2^{1/3} Ai'(0)}{Ai(0)}.$$

En effectuant les changements de variables $u = 2^{1/3}\alpha$ et $\sigma = t/2^{1/3}$, on obtient

$$\int_0^\infty \left(e^{-u\sigma} - 1\right) \mathbf{E} \left[e^{-\sqrt{2}\sigma^{3/2} \int_0^1 e(s)\,ds} \right] \frac{2^{1/3} d\sigma}{\sqrt{2}\pi\sigma^3\sqrt{2}} = \frac{2^{1/3} Ai'(u)}{Ai(u)} - \frac{2^{1/3} Ai'(0)}{Ai(0)}. \tag{7.7}$$

En appliquant l'opérateur $\frac{1}{2^{1/3}} \frac{\partial}{\partial u}$ de part et d'autre de l'égalité, on trouve

$$\begin{aligned} -\int_0^\infty \frac{e^{-u\sigma} \phi(\sqrt{2}\sigma^{3/2})}{2\sqrt{\pi\sigma}}\, d\sigma &= \frac{Ai''}{Ai}(u) - \left(\frac{Ai'}{Ai}\right)^2 (u) \\ &= u - \left(\frac{Ai'}{Ai}\right)^2 (u). \end{aligned}$$

Or on a l'asymptotique

$$\begin{aligned} -\int_0^\infty \frac{e^{-u\sigma}\phi(\sqrt{2}\sigma^{3/2})}{2\sqrt{\pi\sigma}}\,d\sigma &= u - \left(\frac{Ai'(u)}{Ai(u)}\right)^2 \\ &\sim -\frac{1}{\sqrt{u}} + O\left(\frac{1}{u^2}\right), \end{aligned}$$

donc on peut utiliser la formule des résidus pour inverser la transformation de Laplace ci-dessus et calculer ϕ.

Or, la formule des résidus, appliquée à $f(z) = \dfrac{g(z)}{l(z)}$, de pôle multiple a d'ordre n, permet d'affirmer que le résidu B_1 en a est donné par

$$\begin{aligned} (n-1)!B_1 &= \frac{d^{n-1}}{dz^{n-1}}\left[\frac{g(z)(z-a)^n}{l(z)}\right]_{z=a} \\ &= \frac{d^{n-1}}{dz^{n-1}}\left[(z-a)^n f(z)\right]_{z=a}. \end{aligned}$$

Pour $n = 2$, il vient

$$\begin{aligned} B_1 &= \frac{d}{dz}\left[(z-a)^2 f(z)\right]_{z=a} \\ &= -af(a) + \left[(z-a)f'(z)\right]_{z=a}. \end{aligned}$$

Dans le cas qui nous intéresse,

$$\frac{\phi(\sqrt{2}\sigma^{3/2})}{2\sqrt{\pi\sigma}} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{u\sigma}\left[\left(\frac{Ai'}{Ai}\right)^2(u) - u\right]\,du = \sum_1^\infty \sigma e^{a_k\sigma},$$

d'où

$$\begin{aligned} \phi(\sqrt{2}\sigma^{3/2}) &= 2\sqrt{\pi}\sigma^{3/2}\sum_1^\infty e^{a_k\sigma} \\ \phi(z) &= \sqrt{2\pi}z\sum_1^\infty e^{a_k(z/\sqrt{2})^{2/3}}. \end{aligned}$$

□

7.3.2 Fonctionnelle quadratique

Proposition 7.6 *La transformée de Laplace ϕ_{2E} de la distribution de la fonctionnelle quadratique de l'excursion est donnée par*

$$\phi_{2E}(\sigma) = \frac{(\sqrt{2\sigma})^{3/2}}{\sinh(\sqrt{2\sigma})^{3/2}}. \tag{7.8}$$

Démonstration. Dans le cas qui nous intéresse, ϕ_{2B} étant la transformée de Laplace de la distribution de la fonctionnelle quadratique du pont brownien, et étant donné

$$\int_0^\infty e^{-\alpha t}\phi_{2B}\left(\frac{1}{8}t^2\right)\frac{dt}{\sqrt{t}} = \int_0^\infty e^{-\alpha t}\phi_{2B}\left(\left(\frac{t}{\sqrt{8}}\right)^2\right)\frac{dt}{\sqrt{t}} = \int_0^\infty e^{-\alpha\sqrt{8}s}\phi_{2B}\left(s^2\right)8^{1/4}\frac{ds}{\sqrt{s}},$$

on pose

$$g(\alpha) = \frac{8^{-1/4}}{\sqrt{2}} \frac{\Gamma(\alpha/\sqrt{8}+\frac{1}{4})}{\Gamma(\alpha/\sqrt{8}+\frac{3}{4})},$$

pour se placer dans les notations du théorème 7.3.

Si on fait le changement de variable $u = \alpha/\sqrt{8}$, $\sigma = t\sqrt{8}$, $\sigma^3 = t^3 8\sqrt{8}$, l'équation

$$\begin{aligned}\int_0^\infty \left(e^{-\alpha t} - 1\right) \phi_{2E}(t^2) \frac{dt}{\sqrt{2\pi t^3}} &= -g^{-1}(\alpha) + \lim_{\alpha \to 0} g^{-1}(\alpha) \\ &= \sqrt{2} \cdot 8^{1/4} \left(-\frac{\Gamma(\alpha/\sqrt{8}+\frac{3}{4})}{\Gamma(\alpha/\sqrt{8}+\frac{1}{4})} + \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \right)\end{aligned}$$

devient

$$\int_0^\infty \left(e^{-u\sigma} - 1\right) \phi_{2E}\left(\frac{\sigma^2}{8}\right) 8^{1/4} \frac{d\sigma}{\sqrt{2\pi\sigma^3}} = \sqrt{2} \cdot 8^{1/4} \left(-\frac{\Gamma(u+\frac{3}{4})}{\Gamma(u+\frac{1}{4})} + \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \right).$$

En dérivant par rapport à u, on obtient

$$\int_0^\infty e^{-u\sigma} \phi_{2E}\left(\frac{\sigma^2}{8}\right) \frac{d\sigma}{\sqrt{2\pi\sigma}} = \sqrt{2} \left(-\frac{\Gamma'(u+\frac{3}{4})}{\Gamma(u+\frac{1}{4})} + \frac{\Gamma(u+\frac{3}{4})\Gamma'(u+\frac{1}{4})}{\Gamma(u+\frac{1}{4})^2} \right),$$

puis

$$\frac{\phi_{2E}\left(\frac{\sigma^2}{8}\right)}{\sqrt{2\pi\sigma}} = \frac{\sqrt{2}}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{\sigma s} \left(-\frac{\Gamma'(s+\frac{3}{4})}{\Gamma(s+\frac{1}{4})} + \frac{\Gamma(s+\frac{3}{4})\Gamma'(s+\frac{1}{4})}{\Gamma(s+\frac{1}{4})^2} \right) ds.$$

Pour $|\arg(z)| < \pi$, nous avons

$$-\frac{\Gamma'(z+\frac{3}{4})}{\Gamma(z+\frac{1}{4})} + \frac{\Gamma(z+\frac{3}{4})\Gamma'(z+\frac{1}{4})}{\Gamma(z+\frac{1}{4})^2} = -\frac{1}{2z^{1/2}} + \frac{3}{128}\frac{1}{z^{5/2}} + O\left(\frac{1}{z^{7/2}}\right),$$

donc on peut utiliser le théorème des résidus pour calculer l'inversée de la transformée de Laplace.

Comme $\dfrac{\Gamma(z)}{\Gamma(z-\frac{1}{2})}$ a des pôles simples en $z = -n$ pour $n \in \mathbb{N}$, ses dérivées ont des pôles doubles en $z = -n$, si bien que

$$-\frac{\Gamma'(-n+h)}{\Gamma(-n+h-\frac{1}{2})} + \frac{\Gamma(-n+h)\Gamma'(-n+h-\frac{1}{2})}{\Gamma(-n+h-\frac{1}{2})^2} \underset{h\to 0}{=} -\frac{(-1)^n}{n!\Gamma(-n-\frac{1}{2})}\frac{1}{h^2} + O(1).$$

D'autre part,

$$e^{\sigma(-n+h-\frac{3}{4})} = e^{-(n+\frac{3}{4})\sigma} e^{\sigma h} \underset{h\to 0}{=} e^{-(n+\frac{3}{4})\sigma}(1 + \sigma h + O(h^2)),$$

si bien que :

$$\operatorname{Res}\left(e^{\sigma(s-\frac{3}{4})} \left(-\frac{\Gamma'(s+\frac{3}{4})}{\Gamma(s+\frac{1}{4})} + \frac{\Gamma(s+\frac{3}{4})\Gamma'(s+\frac{1}{4})}{\Gamma(s+\frac{1}{4})^2} \right), s = -n \right) = -\frac{(-1)^n}{n!} \frac{\sigma e^{-\frac{3}{4}\sigma - n\sigma}\sigma}{\Gamma(-n-\frac{1}{2})}.$$

Étant donné $\Gamma(-1/2) = -2\sqrt{\pi}$, on trouve donc

$$\begin{aligned}\frac{\phi_{2E}\left(\frac{\sigma^2}{8}\right)}{\sqrt{2\pi\sigma}} &= -\sqrt{2} \sum_{n\in\mathbb{N}} \frac{(-1)^n}{n!} \frac{\sigma e^{-\frac{3}{4}\sigma - n\sigma}}{\Gamma(-n-\frac{1}{2})} = -\sqrt{2} \frac{\sigma e^{-\frac{3}{4}\sigma}}{\Gamma\left(-\frac{1}{2}\right)} \sum_{n\in\mathbb{N}} \binom{-\frac{1}{2}}{n} (e^{-\sigma})^n \\ &= \frac{\sqrt{2}\sigma e^{-\frac{3}{4}\sigma}}{2\sqrt{\pi}} (1 - e^{-\sigma})^{-\frac{3}{2}},\end{aligned}$$

pour finalement obtenir le résultat annoncé

$$\phi_{2E}(\sigma) = \frac{(\sqrt{2\sigma})^{3/2}}{\sinh(\sqrt{2\sigma})^{3/2}}. \tag{7.9}$$

□

7.4 Calcul des moments

Dans cette section, nous montrons qu'il est possible de calculer des moments joints de tout ordre de la distribution de l'intégrale d'une fonction puissance d'un mouvement brownien. Ces calculs restent fastidieux, difficiles à implémenter à cause d'intégrales multiples avec un nombre de plus en plus grand de signes d'intégration.

7.4.1 Cas univarié

Dans le cas de la fonctionnelle aire, les formules (7.2) et (7.7) et le développement asymptotique des fonctions d'Airy permettent d'obtenir des récurrences sur les moments et de calculer ceux-ci.

Dans le cas de la fonctionnelle quadratique, on dispose de formules explicites pour les transformées de Laplace, et un simple développement en série permet de trouver les moments. Notons cependant que le calcul de ces moments directement à partir de la double transformée serait plus difficile que dans le cas de la fonctionnelle précédente, car il faut utiliser l'expansion asymptotique des fonctions Γ.

7.4.2 Moments joints

Le calcul direct d'un moment (joint ou non) d'un ordre choisi est possible. Calculons par exemple le moment joint d'ordre (1,1) de l'aire et du moment d'inertie d'un pont brownien. En procédant comme Mark Kac [121], nous avons :

$$\begin{aligned}
&\mathbf{E}\left[\left(\int_0^\beta |W(\tau)|\, d\tau\right)\left(\int_0^\beta W(\tau)^2\, d\tau\right)\Big| W(\beta)=0\right] \\
&= 2\mathbf{E}\left[\int_0^\beta\int_0^{\tau_2} W(\tau_1)^2|W(\tau_2)|d\tau_1 d\tau_2\Big| W(\beta)=0\right] \\
&= 2\int_0^\beta\int_0^{\tau_2} \mathbf{E}\left[W(\tau_1)^2|W(\tau_2)|\Big| W(\beta)=0\right] d\tau_1 d\tau_2 \\
&= 2\int_0^\beta\int_0^{\tau_2} \lim_{\varepsilon\to 0}\left\{\mathbf{E}\left[W(\tau_1)^2|W(\tau_2)|; -\varepsilon < W(\beta) < \varepsilon\right]/\mathbf{Pr}\left[-\varepsilon < W(\beta) < \varepsilon\right]\right\} d\tau_1 d\tau_2 \\
&= 2\int_0^\beta\int_0^{\tau_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x_2|x_1^2 \frac{e^{-\frac{1}{2}\frac{x_1^2}{\tau_1}}}{\sqrt{2\pi\tau_1}} \frac{e^{-\frac{1}{2}\frac{(x_2-x_1)^2}{\tau_2-\tau_1}}}{\sqrt{2\pi(\tau_2-\tau_1)}} \sqrt{\frac{\beta}{\beta-\tau_2}} e^{-\frac{1}{2}\frac{x_1^2}{\beta-\tau_2}}\, dx_1 dx_2 d\tau_1 d\tau_2,
\end{aligned}$$

car

$$\frac{\sqrt{2\pi\beta}\int_{-\varepsilon}^{\varepsilon} e^{-\frac{1}{2}\frac{(x-u)^2}{(\beta-\tau)}}\, dx}{\sqrt{2\pi(\beta-\tau)}\int_{-\varepsilon}^{\varepsilon} e^{-\frac{x^2}{2\beta}}\, dx} \xrightarrow{\varepsilon\to 0} \frac{\sqrt{\beta}}{\sqrt{\beta-\tau}} e^{-\frac{1}{2}\frac{u^2}{(\beta-u)}}.$$

Il ne reste plus qu'à effectuer le calcul de l'intégrale multiple. En utilisant les formules (pour $a, b, T > 0$) :

$$\begin{aligned}\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}(ax^2-2bx)}\, dx &= \sqrt{2\pi} e^{\frac{b}{2a}} \left(\frac{b^2}{a^{5/2}} + \frac{1}{a^{3/2}}\right) \\ \int_0^{\infty} x e^{-\frac{x^2}{2T}} &= T \\ \int_0^{\infty} x^3 e^{-\frac{x^2}{2T}} &= 2T^2,\end{aligned}$$

on trouve finalement

$$\begin{aligned}\mathbf{E}\left[\left(\int_0^{\beta} |W(\tau)|\, d\tau\right)\left(\int_0^{\beta} W(\tau)^2\, d\tau\right) \middle| W(\beta) = 0\right] &= \sqrt{2\pi}\left(\frac{1}{128} + \frac{5}{768}\right)\beta^{7/2}, \\ \mathbf{E}\left[\left(\int_0^{1} |W(\tau)|\, d\tau\right)\left(\int_0^{1} W(\tau)^2\, d\tau\right) \middle| W(1) = 0\right] &= \frac{11\sqrt{2\pi}}{384}.\end{aligned}$$

Calculons maintenant le moment joint d'ordre (1,1) de l'aire et du moment d'inertie d'une excursion brownienne. Pour cela, en reprenant les notations de [143], on utilise les covariances généralisées K définies dans [143] par

$$K(x_1, x_2, \cdots, x_d) := \mathbf{E}[\tau^+(x_1)\tau^+(x_2)\cdots\tau^+(x_d)],\ x_1 \leqslant x_2 \leqslant \cdots \leqslant x_d,$$

telles que, pour toute fonctionnelle g,

$$\mathbf{E}[G^d] = d! \int_0^{\infty} dx_1 \int_{x_1}^{\infty} dx_2 \cdots \int_{x_{d-1}}^{\infty} dx_d\ g(x_1)g(x_2)\cdots g(x_d)\cdot K(x_1, x_2, \cdots, x_d).$$

Si $a > 0$, et $m_x := \inf\{s \mid X(s) = x\}$, il existe une fonction Θ telle que

$$\mathbf{E}\left[e^{-\sum_1^d \beta_i \tau^+(x_i)},\ 1 > m_{x_d}\right] = \frac{1}{\sqrt{2\pi}i} \int_{a-\infty}^{a+\infty} e^{\alpha}\Theta(d)\, d\alpha, \quad x_d > x_{d-1} > \cdots > x_1,$$

et en posant

$$E(i) := e^{\sqrt{2\alpha}x_i},$$

les transformées de Laplace $\overline{K}$ de K vérifient, pour $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_d$,

$$\overline{K}(\alpha, x_1, \cdots, x_m) = \left|\frac{\partial^m \Theta}{\partial\beta_m \cdots \partial\beta_2 \partial\beta_1}\right|_{\beta_1, \beta_2, \cdots, \beta_m = 0},$$

et en particulier

$$\overline{K}(\alpha, x_1) = \left|\frac{\partial\Theta}{\partial\beta_1}\right|_{\beta_1=0} = 2E(1)^{-2},$$

$$\overline{K}(\alpha, x_1, x_2) = \left|\frac{\partial^2\Theta}{\partial\beta_2\partial\beta_1}\right|_{\beta_1,\beta_2=0} = -\frac{2\sqrt{2}}{\sqrt{\alpha}} E(2)^{-2}\left(E(1)^{-2} - 1\right).$$

Donc, si on définit

$$f_1(\alpha) = \int_0^{\infty}\int_{x_1}^{\infty} \sqrt{\frac{2}{\alpha}} x_2^2 e^{-2\sqrt{2\alpha}x_2}(e^{-2\sqrt{2\alpha}x_1} - 1)\, dx_2\, dx_1 = \frac{85}{512\alpha^3}$$

et

$$f_2(\alpha) = \int_0^\infty \int_{x_1}^\infty \sqrt{\frac{2}{\alpha}} x_2 e^{-2\sqrt{2\alpha}x2}(e^{-2\sqrt{2\alpha}x_1} - 1)\, dx_2\, dx_1 = \frac{59}{512\alpha^3},$$

et si on note F_1 et F_2 leurs inversées de Laplace en α, alors

$$F_1(t) = \frac{85}{1024}\sqrt{2\pi}t^2, \quad F_2(t) = \frac{59}{1024}\sqrt{2\pi}t^2,$$

et le moment d'ordre (1,1) cherché vérifie :

$$\begin{aligned} \mathbf{E}\left[\left(\int_0^1 e(\tau)\, d\tau\right)\left(\int_0^1 e(\tau)^2\, d\tau\right)\right] &= \sqrt{2\pi}\,(F_1(1) + F_2(1)) \\ &= \frac{9}{64}\sqrt{2\pi}. \end{aligned}$$

Un calcul direct de cette dernière valeur est théoriquement possible en utilisant

$$K(x_1, x_2) = \mathbf{E}[\tau^+(x_1)\tau^+(x_2)] = -4[e^{-2(x_1+x_2)^2} - e^{-2x_1^2}],\ x_2 \geqslant x_1,$$

si bien que

$$\mathbf{E}\left[\left(\int_0^1 e(\tau)\, d\tau\right)\left(\int_0^1 e(\tau)^2\, d\tau\right)\right] = \int_0^\infty dx_1 \int_{x_1}^\infty dx_2\, x_1 x_2^2 \big(K(x_1, x_2) + K(x_2, x_1)\big).$$

7.5 Conclusion

À l'aide de la formule de Feynman-Kac, du théorème de Louchard et de manipulations relativement élémentaires, nous avons obtenu des expressions pour la transformée de Laplace de la distribution des fonctionnelles aire et moment d'inertie des pont et excursion browniens, ainsi qu'une expression pour la double transformée de Laplace de la distribution jointe de ces deux fonctionnelles.

Alors que le calcul des moments de ces deux fontionnelles se faisait assez facilement, il apparaît difficile de dégager un schéma de récurrence pour calculer les moments joints. En effet, il n'existe pas de développement asymptotique disponible en $+\infty$ pour le membre de droite de la formule (7.5).

Nous verrons dans le chapitre suivant qu'un pompage de moments sur des séries génératrices permet de trouver un tel schéma.

Chapitre 8

Pompage de moments

Dans ce chapitre, nous étudions l'aire et la somme des carrés des hauteurs (aussi appelée fonctionnelle quadratique ou moment d'inertie) des chemins de Dyck et des chemins de Dyck bilatères, aussi appelés chemins de Grand Dyck. Nous utilisons pour cela la méthode de pompage des moments, qu'on peut appeler plus simplement méthode des moments, à base de séries génératrices développée par Takács [198, 199, 200].

Notons qu'il est bien connu que les chemins de Dyck sont en bijection avec les arbres de Catalan — voir par exemple les papiers de Takács mentionnés ci-dessus, voire [193] pour trouver une liste d'objets combinatoires énumérés par les nombres de Catalan —, et qu'étudier la somme d'une fonction des hauteurs sur les chemins de Dyck revient à étudier la somme d'une fonction des distances des nœuds à la racine d'un arbre binaire. Une correspondance bijective entre chemin de Dyck et arbre binaire peut se résumer en quelques mots : les pas du chemin de Dyck correspondent aux arêtes de l'arbre parcouru dans l'ordre préfixe. Ainsi, l'aire sous un chemin de Dyck correspond à la longueur de cheminement interne gauche d'un arbre binaire, et le moment d'inertie d'un chemin de Dyck par rapport à l'axe des abscisses correspond au moment d'inertie (en ne prenant que les arêtes gauches en compte dans le calcul des distances des nœuds à la racine) d'un arbre binaire par rapport à sa racine.

Pour une fonctionnelle de l'ensemble des distances des nœuds à la racine, il importe peu que ces distances soient calculées en prenant en compte les arêtes gauches et/ou droites, ou qu'on prenne en compte les nœuds internes et/ou externes : on obtient des calculs similaires et les mêmes distributions limites.

On parle de pompage car pour calculer le moment d'un ordre donné r désiré, on dérive r fois une certaine équation fonctionnelle pour en tirer, après quelques manipulations, une récurrence entre ce moment et les moments d'ordre strictement plus petits que r. C'est cette opération de dérivation, à effectuer plusieurs fois, qu'on appelle un peu familièrement, pompage.

C'est la première fois à notre connaissance qu'on utilise ainsi le méthode des moments pour des paramètres multivariés. Nous avons vu dans le chapitre précédent que le calcul de moments joints de l'aire et du moment d'inertie du mouvement brownien, en particulier du pont et de l'excursion, pouvait s'avérer ardu. Or la méthode des moments nous permet au passage de trouver des récurrences faciles à implémenter pour calculer les moments joints de la distribution limite de ces deux paramètres.

En particulier, si b désigne le pont brownien normalisé et e l'excursion brownienne normalisée, nous montrons qu'on a des moments joints sous la forme

$$\mathbf{E}\left[\left(\int_0^1 |b(t)|\,dt\right)^r \left(\int_0^1 b(t)^2\,dt\right)^s\right] = \frac{B_{r,s}}{2^{\frac{3r}{2}+2s}} \frac{\sqrt{\pi} r!s!}{\Gamma\left(\frac{3r+4s+1}{2}\right)}$$

et

$$\mathbf{E}\left[\left(\int_0^1 e(t)\,dt\right)^r \left(\int_0^1 e(t)^2\,dt\right)^s\right] = \frac{4A_{r,s}}{2^{\frac{3r}{2}+2s}} \frac{\sqrt{\pi} r! s!}{\Gamma\left(\frac{3r+4s-1}{2}\right)},$$

avec certaines constantes rationnelles $B_{r,s}$ et $A_{r,s}$ vérifiant certaines relations de récurrence.

Déterminer ces récurrences, néanmoins, peut se révéler difficile, mais le fait que les moments de l'aire et du moment d'inertie étaient déjà connus permet de valider *a posteriori* ces calculs, car dans nos récurrences les moments de la fonctionnelle quadratique dépendent des moments joints. Le mécanisme du pompage de moments bivarié étant bien débrousaillé, il serait maintenant aisé de le généraliser aux paramètres somme-de-puissances pour une puissance fixée.

Nous étendons cette étude aux méandres (chemins partant de zéro et restant au-dessus de l'axe des abscisses) et aux marches de Bernoulli quelconques dont nous retrouvons la distribution de l'aire. Nous montrons en outre que cette méthodologie par fonctions génératrices permet de retrouver des relations entre doubles transformées de Laplace de fonctionnelles polynomiales de divers mouvements browniens, qui, si elles étaient déjà connues des probabilistes, n'en restent pas moins assez élaborées.

Notons que la méthode des moments permet également de traiter les fonctionnelles signées et permet de retrouver des relations entre doubles transformées de Laplace de fonctionnelles signées et non signées.

Même si cette méthode semble faire double emploi avec le mouvement brownien, elle permet tout de même de retrouver des lois de manière constructive. En outre, cette méthode est inévitable dans le cas où le contour de l'arbre ne converge pas forcément vers une excursion brownienne, comme dans le cas des arbres non orientés. Le fait de connaître la distribution limite du moment d'inertie à partir du mouvement brownien, des récurrences sur les moments de cette distribution qui l'identifie de manière unique, et de retrouver ces récurrences sur d'autres classes d'arbres, permet d'affirmer que la distribution limite du moment d'inertie de ces classes d'arbres est la même et de l'identifier.

8.1 Convergence des moments

Drmota [47, 48] a récemment montré la convergence des moments pour les fonctionnelles polynomialement bornées de marches discrètes convergeant vers un mouvement brownien, sans avoir recours au théorème de Carleman. Auparavant, seule la convergence en loi était connue *a priori* [15].

Théorème 8.1 *Supposons qu'une distribution de descendants ξ d'un processus de Galton-Watson est critique, i.e.* $\mathbf{E}[\xi] = 1$, *et que la variance* $\sigma^2 = \mathbf{Var}[\xi]$ *soit non nulle et finie. Alors le processus de recherche en profondeur rééchelonné converge faiblement vers* $\frac{\sqrt{2}}{\sigma}e(t)$, *où e désigne l'excursion brownienne standard de durée 1 :*

$$\left(\frac{1}{\sqrt{2n}} X_n(2nt),\ 0 \le t \le 1\right) \xrightarrow{\mathcal{D}} \left(\frac{\sqrt{2}}{\sigma} e(t),\ 0 \le t \le 1\right).$$

Théorème 8.2 *Supposons qu'une suite de processus stochastiques $x_n(t)$ définie sur $\mathcal{C}[0,1]$ converge faiblement vers $x(t)$. Supposons de plus qu'il existe $s_0 \in [0,1]$ tel que pour tout $r \ge 0$,*

$$\sup_{n \ge 0} \mathbf{E}\left[|x_n(s_0)|^r\right] < \infty,$$

et que, pour tout $\alpha > 1$, il existe $\beta > 0$ et $C > 0$ avec

$$\mathbf{E}\left[|x_n(t) - x_n(s)|^\beta\right] \leq C|t-s|^\alpha, \quad \forall s,t \in [0,1].$$

Soit $F : C[0,1] \to \mathbb{R}$ une fonctionnelle continue à croissance polynomiale, i.e. telle qu'il existe $r \geq 0$ avec

$$|F(y)| \leq \|y\|_\infty^r$$

pour tout $y \in C[0,1]$. Alors

$$\lim_{n\to\infty} \mathbf{E}F(x_n) = \mathbf{E}F(x).$$

8.2 Aire et moment d'inertie joints sous les chemins de Grand Dyck

8.2.1 Équation fonctionnelle

Proposition 8.1 *Soit $D(z,u,v)$ la fonction génératrice trivariée énumérant les chemins de Dyck selon leur longueur (variable z), l'aire S_1 (variable u), et la fonctionnelle quadratique S_2 (variable v). Alors $D(z,u,v)$ satisfait :*

$$D(z,u,v) = 1 + z^2uvD(z,u,v)D(zuv,uv^2,v).$$

Démonstration. Considérons une arche positive a de longueur $2n+2$, et soient $h_0, h_1, \cdots, h_{2n+2}$ les altitudes de ses points. les entiers $h'_1 = h_1 - 1, h'_2 = h_2 - 1, \cdots, h'_{2n+1} = h_{2n+1} - 1$ définissent une chemin de Dyck d, et de

$$\sum_{i=0}^{2n+2} h_i^2 = \sum_{i=1}^{2n+1} (1+h'_i)^2$$

on déduit $S_1(a) = 2n+1+S_1(d)$ et $S_2(a) = 2n+1+2S_1(d)+S_2(d)$. La fonction génératrice des arches positives est donc donnée par $z^2uvD(zuv,uv^2,v)$. Comme un chemin de Dyck est soit un chemin vide, soit un chemin de Dyck suivie d'une arche positive, nous avons l'équation fonctionnelle voulue. □

8.2.2 Pompage des moments

Afin de simplifier les calculs, nous introduisons la fonction $G(z,u,v) = zD(z,u,v)$, qui vérifie

$$G(z,u,v) = z + zG(z,u,v)G(zuv,uv^2,v).$$

Alors nous avons la proposition :

Proposition 8.2 *Soit $\eta_{r,s}$ le moment d'ordre (r,s) de la distribution jointe de l'aire et de la fonctionnelle quadratique de l'excursion. Autrement dit, si $X = \int_0^1 e(s)\,ds$ et $Y = \int_0^1 e^2(s)\,ds$, où e est une excursion normalisée, $\eta_{r,s} = \mathbf{E}[X^rY^s]$. Alors il existe des constantes $A_{r,s} \neq 0$ telles que, pour tous r,s,*

$$\eta_{r,s} = \frac{4A_{r,s}}{2^{\frac{3r}{2}+2s}} \frac{\sqrt{\pi}r!s!}{\Gamma\left(\frac{3r+4s-1}{2}\right)}.$$

Elles peuvent être calculées à l'aides des récurrences suivantes :

$$\begin{cases} A_{0,0} &= -\frac{1}{2}, \\ A_{r,s} &= (r+1)A_{r+1,s-1} + \sum_{j=0}^{s}\sum_{i=0}^{r} \mathbb{I}_{[i+j>0, r+s>i+j]} A_{i,j}A_{r-i,s-j} \\ & \quad + \mathbb{I}_{r>0}\frac{(3r+4s-4)}{2}A_{r-1,s}. \end{cases} \tag{8.2}$$

Avant de procéder à la preuve de la proposition elle-même, nous allons montrer un résultat intermédiaire matérialisé par le lemme 8.1 en-dessous, pour lequel nous avons besoin d'introduire la notion de degré de série de Laurent :

Définition 8.1 *Soit f une série de Laurent en x, comprenant un nombre fini de termes en x^n pour $n \in \mathbb{N}$. On appellera degré de f le degré du polynôme formé des termes en x^n de f.*

Lemme 8.1 *Pour tous $r, s \in \mathbb{N}$ tels que $r + s > 0$, $zG_{u^r v^s}(z, 1, 1)$ est une série de Laurent en $1/R$ de degré fini, égal à $3r + 4s - 1$, où $R = \sqrt{1 - 4z^2}$. Pour $r = s = 0$, nous avons $zG(z, 1, 1) = \frac{1-R}{2}$. Si nous définissons $(A_{r,s})$ de telle manière que*

$$\frac{z}{r!s!} G_{u^r v^s}(z, 1, 1) \underset{z \to \frac{1}{2}}{\sim} A_{r,s} / R^{3r+4s-1},$$

alors $(A_{r,s})$ est donnée par la récurrence (8.2).

Démonstration. Notons que, excepté pour $r = s = 0$ pour lesquels $zG(z, 1, 1) = \frac{1-R}{2}$, $A_{r,s}$ est le coefficient dominant de $\frac{z}{r!s!} G_{u^r v^s}(z, 1, 1)$, vu comme une série de Laurent en $1/R$.

On montre le lemme par récurrence on (r, s), où $(r', s') < (r, s)$ si $r' + s' < r + s$ ou $(r' + s' = r + s$ and $s' < s)$. la propriété voulue est vraie pour $r = s = 0$. Maintenant fixons r, s de telle sorte que $r + s > 0$ et supposons le résultat vrai pour r', s' tel que $(r', s') < (r, s)$.

Alors, la formule

$$\frac{\partial^{r+s} G(z, u, v)}{\partial u^r v^s} = \delta_{r0}\delta_{s0} z + z \sum_{j=0}^{s} \binom{s}{j} \sum_{i=0}^{r} \binom{r}{i} \frac{\partial^{i+j} G(z, u, v)}{\partial u^i \partial v^j} \frac{\partial^{r-i+s-j} G(zuv, uv^2, v)}{\partial u^{r-i} \partial v^{s-j}}$$

entraîne

$$\begin{aligned}
& (1 - 2zG(z, 1, 1)) z \frac{\partial^{r+s} G(z, u, v)}{r!s! \partial u^r \partial v^s}\Big|_{u=v=1} \\
= \; & z \sum_{\substack{0 \leqslant i \leqslant r, 0 \leqslant j \leqslant s \\ 0 < i+j < r+s}} \binom{s}{j} \binom{r}{i} z \frac{\partial^{i+j} G(z, u, v)}{\partial u^i \partial v^j}\Big|_{u=v=1} \cdot z \frac{\partial^{r-i+s-j} G(zuv, uv^2, v)}{\partial u^{r-i} \partial v^{s-j}}\Big|_{u=v=1} \\
& + zG(z, 1, 1) \left(z \frac{\partial^{r+s} G(zuv, uv^2, v)}{\partial u^r \partial v^s}\Big|_{u=v=1} - z \frac{\partial^{r+s} G(z, u, v)}{r!s! \partial u^r \partial v^s}\Big|_{u=v=1} \right) \qquad (8.1)
\end{aligned}$$

avec $1 - 2zG(z, 1, 1) = R$. Pour tous i, j tels que $0 < i + j < r + s$,

$$z \frac{\partial^{r+s} G(z, u, v)}{r!s! \partial u^r \partial v^s}\Big|_{u=v=1}$$

est une série de Laurent de degré $3i + 4j - 1$ en $1/R$. Montrons qu'il en est de même pour

$$z \frac{\partial^{r+s} G(zuv, uv^2, v)}{\partial u^{r-i} \partial v^{s-j}}\Big|_{u=v=1}.$$

En utilisant la formule de Faa di Bruno [160], nous obtenons, pour $i + j > 0$,

$$\begin{aligned}
& z \frac{\partial^j G(zuv, uv^2, v)}{\partial v^j} \\
= \; & \sum_{k_1 + 2k_2 = j} \frac{j!}{k_2!} \sum_{q_1 + q_2 + q_3 = k_1} \frac{2^{q_2}}{q_1! q_2! q_3!} \cdot z^{q_1} u^{q_1 + q_2 + k_2} G_{z^{q_1} u^{q_2 + k_2} v^{q_3}}(zuv, uv^2, v),
\end{aligned}$$

donc

$$\begin{aligned}
&z\frac{\partial^{i+j}G(zuv,uv^2,v)}{\partial u^i\partial v^j}\\
&= \frac{\partial^i}{\partial u^i}\sum_{k_1+2k_2=j}\frac{j!}{k_2!}\sum_{q_1+q_2+q_3=k_1}\frac{2^{q_2}}{q_1!q_2!q_3!}\cdot z^{q_1}u^{q_1+q_2+k_2}G_{z^{q_1}u^{q_2+k_2}v^{q_3}}(zuv,uv^2,v)\\
&= \sum_{k_1+2k_2=j}\frac{j!}{k_2!}\sum_{q_1+q_2+q_3=k_1}\frac{2^{q_2}}{q_1!q_2!q_3!}\cdot z^{q_1}\frac{\partial^i}{\partial u^i}u^{q_1+q_2+k_2}G_{z^{q_1}u^{q_2+k_2}v^{q_3}}(zuv,uv^2,v). \qquad (8.2)
\end{aligned}$$

Il est clair que si zF est une série de Laurent quelconque en $1/R$ de degré $p\neq 0$, alors zF_z est une série de Laurent de degré $p+2n$ en $1/R$. Il est alors clair que, pour $n\geqslant 1$, $zG_{z^n}(z,1,1)$ une série de Laurent de degré $2n-1$ en $1/R$. En outre, les termes dans la formule (8.2)

$$\frac{\partial^i}{\partial u^i}u^{q_1+q_2+k_2}G_{z^{q_1}u^{q_2+k_2}v^{q_3}}(zuv,uv^2,v)\Big|_{u=v=1}$$

sont tous des séries de Laurent de degré $2q_1+3(i+q_2+k_2)+4q_3-1$ en $1/R$ (sauf éventuellement $G_{u^rv^s}(z,1,1)$ qui reste inconnu).

Nous en déduisons en premier lieu que dans la formule (8.1), pour $0<i+j<r+s$, les termes

$$z\frac{\partial^{i+j}G(z,u,v)}{\partial u^i\partial v^j}\Big|_{u=v=1}$$

et

$$z\frac{\partial^{r-i+s-j}G(zuv,uv^2,v)}{\partial u^{r-i}\partial v^{s-j}}\Big|_{u=v=1}$$

sont des séries de Laurent de degrés respectifs $3i+4j-1$ et $3(r-i)+4(s-j)-1$, si bien que leurs produits sont des séries de Laurent de degré $3i+4j-1+3(r-i)+4(s-i)-1=3r+4s-2$.

Nous en déduisons en second lieu, à partir de la formule (8.2) appliquée à $(i,j)=(r,s)$, que

$$z\frac{\partial^{r+s}G(zuv,uv^2,v)}{\partial u^r\partial v^s}\Big|_{u=v=1}-z\frac{\partial^{r+s}G(z,u,v)}{r!s!\partial u^r\partial v^s}\Big|_{u=v=1}$$

est une somme de termes de la forme $G_{z^iu^jv^k}(z,1,1)$ avec $i+j+k\leqslant r+s$ et $(j,k)<(r,s)$, et donc est une série de Laurent de degré strictement plus petit que $3r+4s-1$ et est en fait une série de Laurent de degré $3r+4s-2$, à cause de la présence des termes $G_{u^{r+1}v^{s-1}}$ et $G_{zu^{r-1}v^s}$ (ce dernier terme vient de $\frac{\partial^r}{\partial u^r}uG_{v^s}(zuv,uv^2,v)$).

Avec ces résultats en tête, nous déduisons finalement, de la formule (8.1) que $zG_{u^rv^s}(z,1,1)$ est une série de Laurent en $1/R$ de degré $3r+4s-1$. Nous pouvons définir $r!s!A_{r,s}$ comme étant son coefficient dominant.

Calculons le coefficient de plus grand degré en $1/R$ de

$$\begin{aligned}
&z\frac{\partial^{r+s}G(zuv,uv^2,v)}{\partial u^r\partial v^s}\Big|_{u=v=1}-z\frac{\partial^{r+s}G(z,u,v)}{r!s!\partial u^r\partial v^s}\Big|_{u=v=1}\\
&= \sum_{k_1+2k_2=s}\frac{s!}{k_2!}\sum_{q_1+q_2+q_3=k_1}\frac{2^{q_2}}{q_1!q_2!q_3!}\cdot z^{q_1}\frac{\partial^r}{\partial u^r}u^{q_1+q_2+k_2}G_{z^{q_1}u^{q_2+k_2}v^{q_3}}(zuv,uv^2,v)\Big|_{u=v=1}\\
&\quad -zG_{u^rv^s}(z,1,1)\\
&= \sum_{\substack{k_1+2k_2=s,q_1+q_2+q_3=k_1\\(k_1,k_2,q_1,q_2,q_3)\neq(s,0,0,0,s)}}\frac{s!}{k_2!}\frac{2^{q_2}}{q_1!q_2!q_3!}\cdot z^{q_1}\frac{\partial^r}{\partial u^r}u^{q_1+q_2+k_2}G_{z^{q_1}u^{q_2+k_2}v^{q_3}}(zuv,uv^2,v)\Big|_{u=v=1}.
\end{aligned}$$

Il proviennent des termes ci-dessus en $1/R^{3r+4s-2}$ pour $(k_1, k_2, q_1, q_2, q_3)$ égal à $(s, 0, 0, 1, s-1)$ et $(s, 0, 1, 0, s-1)$. En utilisant

$$\frac{z}{(r-1)!s!} G_{zu^{r-1}v^s}(z,1,1) \underset{z\to\frac{1}{2}}{\sim} 2(3r+4s-2)A_{r-1,s+1}/R^{3r+4s-2},$$

obtenu par dérivation par rapport à z de $G_{u^{r-1}v^s}(z,1,1)$, faisant notamment intervenir le calcul

$$\frac{d}{dz}\frac{1}{R^{3(r-1)+4s-1}} = \frac{d}{dz}\frac{1}{(1-4z^2)^{\frac{3r+4s-4}{2}}} = \frac{8z\cdot\frac{3r+4s-2}{2}}{(1-4z^2)^{\frac{3r+4s-2}{2}}} = \frac{4z(3r+4s-2)}{R^{3r+4s-2}},$$

nous trouvons

$$\begin{aligned}
& z\frac{\partial^{r+s}G(zuv,uv^2,v)}{r!s!\partial u^r\partial v^s}\Big|_{u=v=1} - zG_{u^rv^s}(z,1,1) \\
= \;& 2s\frac{(r+1)!(s-1)!}{r!s!}A_{r+1,s-1}/R^{3r+4s-2} + r\frac{(r-1)!s!}{r!s!}\cdot\frac{1}{2}\cdot 2(3r+4s-4)A_{r-1,s}/R^{3r+4s-2}+\cdots \\
= \;& 2(r+1)A_{r+1,s-1}/R^{3r+4s-2} + (3r+4s-4)A_{r-1,s}/R^{3r+4s-2}+\cdots,
\end{aligned}$$

où les termes restants représentent une série de Laurent en $1/R$ de degré strictement plus petit que $3r+4s-2$.

Finalement, par identification dans (8.1)

$$\begin{aligned}
A_{r,s} &= \frac{1}{2}\cdot 2(r+1)A_{r+1,s-1} + \sum_{j=0}^{s}\sum_{i=0}^{r}\mathbb{I}_{[i+j>0,r+s>i+j]}A_{i,j}A_{r-i,s-j} \\
&\quad + \mathbb{I}_{r>0}\frac{1}{2}\cdot 2(3r+4s-4)A_{r-1,s} \\
&= (r+1)A_{r+1,s-1} + \sum_{j=0}^{s}\sum_{i=0}^{r}\mathbb{I}_{[i+j>0,r+s>i+j]}A_{i,j}A_{r-i,s-j} \\
&\quad + \mathbb{I}_{r>0}\frac{(3r+4s-4)}{2}A_{r-1,s},
\end{aligned}$$

ce qui constitue la formule annoncée. □

Preuve de la Proposition 8.2. Comme $[z^n]G_{u^pv^q}(z,1,1)$ (le coefficient de z^n dans le développement en série de $G_{u^pv^q}(z,1,1)$) est négligeable devant $[z^n]G_{u^rv^s}(z,1,1)$ dès que $p+q<r+s$, les moments factoriels sont les mêmes que les moments eux-mêmes. En utilisant l'asymptotique bien connue des nombres de Catalan [81], on obtient le résultat voulu. □

Il est maintenant facile de calculer les moments joints avec quelques lignes de Maple, ce qui mène aux tableaux 7.1 and 7.2, où nous avons mis les premières valeurs de $A_{r,s}$ et $\eta_{r,s}$. Pour $r=0$ et $s=0$ nous retrouvons dans le tableau 2 les valeurs trouvées respectivement par Takács [198] ou en effectuant le développement en série de ϕ_{2B}.

8.2.3 Moments joints de chemins de Grand Dyck

Prouvons maintenant des résultats similaires pour le pont brownien. Nous allons maintenant diriger notre attention sur les chemins de Grand Dyck.

Proposition 8.3 *Soit $\beta_{r,s}$ le moment d'ordre (r,s) de la distribution jointe de l'aire et du moment d'inertie de l'excursion. En d'autres termes, si $X=\int_0^1 b(s)\,ds$ et $Y=\int_0^1 b^2(s)\,ds$, alors*

TAB. 8.1 – Valeurs de $A_{r,s}$.

r,s	0	1	2	3	4	5
0	-1/2	12	13/4	63	37297/16	280371/2
1	1/4	3/2	239/8	4515/4	2199119/32	24733149/4
2	5/16	107/16	8461/32	66045/4	192844087/128	3045954135/16
3	15/16	159/4	2585	482475/2	3951804863/128	83537034105/16
4	$\frac{1105}{256}$	$\frac{75783}{256}$	$\frac{14599815}{512}$	$\frac{475670055}{128}$	$\frac{1300934713135}{2048}$	$\frac{35461554388785}{256}$
5	$\frac{1695}{64}$	$\frac{84843}{32}$	$\frac{11323097}{32}$	$\frac{3926382201}{64}$	$\frac{3455175997157}{256}$	$\frac{235919056678557}{64}$

TAB. 8.2 – Moments joints de l'aire et du moment d'inertie de l'excursion.

r,s	0	1	2	3	4	5
0	1	1/2	$\frac{19}{60}$	$\frac{631}{2520}$	$\frac{1219}{5040}$	$\frac{92723}{332640}$
1	$\frac{\sqrt{2\pi}}{4}$	$\frac{9\sqrt{2\pi}}{64}$	$\frac{51\sqrt{2\pi}}{512}$	$\frac{7143\sqrt{2\pi}}{81920}$	$\frac{422781\sqrt{2\pi}}{4587520}$	$\frac{25311493\sqrt{2\pi}}{220200960}$
2	$\frac{5}{12}$	$\frac{221}{840}$	$\frac{1643}{7920}$	$\frac{4321753}{21621600}$	$\frac{291886913}{1269777600}$	$\frac{881725807}{2855588736}$
3	$\frac{15\sqrt{2\pi}}{128}$	$\frac{339\sqrt{2\pi}}{4096}$	$\frac{16573\sqrt{2\pi}}{229376}$	$\frac{838799\sqrt{2\pi}}{11010048}$	$\frac{13125751\sqrt{2\pi}}{138412032}$	$\frac{330256199219\sqrt{2\pi}}{2418335023104}$
4	$\frac{221}{1008}$	$\frac{16565}{96096}$	$\frac{40594139}{245044800}$	$\frac{5310071279}{27935107200}$	$\frac{181927010953}{713897184000}$	$\frac{87609843192007}{223592598028800}$
5	$\frac{565\sqrt{2\pi}}{8192}$	$\frac{3935\sqrt{2\pi}}{65536}$	$\frac{463551\sqrt{2\pi}}{7340032}$	$\frac{3648007367\sqrt{2\pi}}{46506442752}$	$\frac{3817396540159\sqrt{2\pi}}{33856690323456}$	$\frac{199793478812293\sqrt{2\pi}}{1083414090350592}$

b est une excursion normalisée, $\beta_{r,s} = \mathbf{E}[X^r Y^s]$. Alors il existe des constantes $B_{r,s} \neq 0$ telles que, pour tous r, s,

$$\beta_{r,s} = \frac{B_{r,s}}{2^{\frac{3r}{2}+2s}} \frac{\sqrt{\pi} r! s!}{\Gamma\left(\frac{3r+4s+1}{2}\right)}.$$

Elles peuvent être calculées à l'aide des récurrences

$$\begin{cases} B_{0,0} &= -\frac{1}{2}, \\ B_{r,s} &= (r+1)B_{r+1,s-1} + \sum_{j=0}^{s}\sum_{i=0}^{r} \mathbb{I}_{[i+j>0, r+s>i+j]} B_{i,j} B_{r-i,s-j} \\ & \quad + \mathbb{I}_{r>0} \frac{(3r+4s-4)}{2} B_{r-1,s}. \end{cases} \tag{8.3}$$

Démonstration. Introduisons la fonction génératrice $P(z,u,v)$ des chemins de Grand Dyck non vides. Un chemin de Grand Dyck non vide est une suite d'arches positives, d'où

$$(1 - 2z^2 uv D(zuv, uv^2, v))P(z,u,v) = 1.$$

Nous en déduisons que, pour tous r, s, $P_{u^r v^s}(z,u,v)$ est une série de Laurent en $1/R$ de degré $3r+4s+1$. Définissons la double suite $(B_{r,s})$ telle que

$$P_{u^r v^s}(z,1,1) \sim r! s! B_{r,s} / R^{3r+4s+1}.$$

Définissons maintenant $H(z,u,v) := 1 - 2z^2 uv D(zuv, uv^2, v) = 1 - 2zG(zuv, uv^2, v)$. De $HP = 1$, nous déduisons

$$\sum_{j=0}^{s}\sum_{i=0}^{r} \binom{s}{j}\binom{r}{i} H_{u^i v^j} P_{u^{r-i} v^{s-j}} = \delta_{r0}\delta_{s0}.$$

En identifiant les coefficients, on obtient

$$\sum_{j=0}^{s}\sum_{i=0}^{r} A_{i,j}B_{r-i,s-j} = \delta_{r0}\delta_{s0}A_{0,0}B_{0,0},$$

ce qui donne déjà une formule de récurrence assez pratique pour calculer les $B_{r,s}$, qui s'écrit

$$A_{0,0} = -1/2, \quad B_{0,0} = 1, \quad B_{r,s} = 2\sum_{j=0}^{s}\sum_{i=0}^{r} \mathbb{I}_{i+j>0,r+s>i+j}A_{i,j}B_{r-i,s-j}.$$

Néanmoins, nous pouvons en fait trouver une récurrence dépendant seulement des $B_{r,s}$ de la manière suivante.

En premier lieu, si nous définissons $A(u,v) = \sum A_{r,s}(-1)^{r+s}u^r v^s$, la récurrence sur les $A_{r,s}$ se traduit par

$$vA_u + A^2 - A(0,0)^2 - \frac{3}{2}u^2A_u + 2uvA_v + \frac{u}{2}A = 0. \quad (8.2.3)$$

De $HP = 1$, nous déduisons que, si nous définissons $B(u,v) = \sum B_{r,s}(-1)^{r+s}u^r v^s$, alors BA est constant, égal à $BA = (BA)(0,0) = -1/2$, d'où $A = -1/(2B)$, $A_u = B_u/(2B^2)$, $A_v = B_v/(2B^2)$, et la formule (8.2.3) devient

$$v\frac{B_u}{2B^2} + \frac{1}{4B^2} - \frac{1}{4} - \frac{3}{2}u^2\frac{B_u}{2B^2} + uv\frac{B_v}{B^2} - \frac{u}{4B} = 0$$

ou

$$2vB_u + 1 - B^2 - 3u^2B_u + 4uvB_v - uB = 0. \quad (8.3)$$

En utilisant

$$\begin{aligned} B(u,v) &= \sum(-1)^{r+s}B_{rs}u^r v^s \\ vB_u &= \sum(-1)^{r+s}rB_{rs}u^{r-1}v^{s+1} \\ u^2B_u &= \sum(-1)^{r+s}rB_{rs}u^{r+1}v^s \\ uvB_v &= \sum(-1)^{r+s}sB_{rs}u^{r+1}v^s \\ uB &= \sum(-1)^{r+s}sB_{rs}u^{r+1}v^s. \end{aligned}$$

nous obtenons, pour $r + s > 0$,

$$2(r+1)B_{r+1,s-1} - \sum_{i=0}^{r}\sum_{j=0}^{s} B_{i,j}B_{r-i,s-j} + (3r+4s-2)B_{r-1,s} = 0.$$

À nouveau, une analyse de singularité [81] nous donne $\beta_{r,s} = \dfrac{B_{r,s}}{2^{\frac{3r}{2}+2s}}\dfrac{\sqrt{\pi}r!s!}{\Gamma\left(\frac{3r+4s+1}{2}\right)}$. □

Nous pouvons facilement calculer les moments $\beta_{r,s}$, dont le lecteur trouvera les premières valeurs dans le tableau 8.4.Pour $r = 0$ et $s = 0$, nous retrouvons les valeurs trouvées par Shepp [190] ou Tolmatz [202, 204, 205].

TAB. 8.3 – Valeurs de $B_{r,s}$.

r,s	0	1	2	3	4	5
0	1	1/2	21/8	671/16	180323/128	20898423/256
1	1/2	11/4	173/4	22931/16	1319183/16	233526463/32
2	$\frac{7}{8}$	$\frac{273}{16}$	$\frac{38733}{64}$	$\frac{4615949}{128}$	$\frac{3344400297}{1024}$	$\frac{856393361595}{2048}$
3	$\frac{21}{8}$	$\frac{2037}{16}$	$\frac{275373}{32}$	$\frac{1672797}{2}$	$\frac{7175989065}{64}$	$\frac{5097582269451}{256}$
4	$\frac{1463}{128}$	$\frac{289555}{256}$	$\frac{133182175}{1024}$	$\frac{39357656121}{2048}$	$\frac{59615450866005}{16384}$	$\frac{28408170184726485}{32768}$
5	$\frac{2121}{32}$	$\frac{187677}{16}$	$\frac{67703769}{32}$	$\frac{57697769535}{128}$	$\frac{59486822349255}{512}$	$\frac{2313457431011811}{64}$

TAB. 8.4 – Moments joints de l'aire et du moment d'inertie du pont brownien.

r,s	0	1	2	3	4	5
0	1	1/6	1/20	$\frac{61}{2520}$	$\frac{1261}{75600}$	$\frac{79}{5280}$
1	$\frac{\sqrt{2\pi}}{8}$	$\frac{11\sqrt{2\pi}}{384}$	$\frac{173\sqrt{2\pi}}{15360}$	$\frac{22931\sqrt{2\pi}}{3440640}$	$\frac{1319183\sqrt{2\pi}}{247726080}$	$\frac{233526463\sqrt{2\pi}}{43599790080}$
2	$\frac{7}{60}$	$\frac{13}{360}$	$\frac{12911}{720720}$	$\frac{355073}{28274400}$	$\frac{53085719}{4655851200}$	$\frac{19030963591}{1499184086400}$
3	$\frac{21\sqrt{2\pi}}{1024}$	$\frac{679\sqrt{2\pi}}{81920}$	$\frac{13113\sqrt{2\pi}}{2621440}$	$\frac{79657\sqrt{2\pi}}{19660800}$	$\frac{68342753\sqrt{2\pi}}{16609443840}$	$\frac{26971334759\sqrt{2\pi}}{5374077829120}$
4	$\frac{19}{720}$	$\frac{8273}{617760}$	$\frac{761041}{79814592}$	$\frac{1874174101}{214169155200}$	$\frac{3974363391067}{404779703328000}$	$\frac{1893878012315099}{145558781316748800}$
5	$\frac{101\sqrt{2\pi}}{16384}$	$\frac{993\sqrt{2\pi}}{262144}$	$\frac{358221\sqrt{2\pi}}{115343360}$	$\frac{549502567\sqrt{2\pi}}{172738215936}$	$\frac{62949018359\sqrt{2\pi}}{16122233487360}$	$\frac{12240515507999\sqrt{2\pi}}{2192623754280960}$

Il semble difficile de trouver une expression sympathique de B (ou A). Cependant, si nous définissons

$$g(\alpha,a,b)=\int_0^\infty e^{-\alpha t}\phi_{12B}\left(at^{3/2},b^2t^2\right)\,\frac{dt}{2\pi t},$$

un calcul direct donne

$$g(\alpha,a,b)=\frac{1}{\sqrt{2\alpha}}B\left(\frac{a}{(2\alpha)^{3/2}},\frac{b^2}{(2\alpha)^2}\right),$$

ou

$$B(u,v)=g(1/2,u,\sqrt{v}),$$

et la formule (7.5) où le chapitre précédent donne l'expression suivante de B :

$$B(u,v)=-\frac{1}{(8v)^{1/4}}\frac{D_{u^2v^{-3/2}2^{-1/2}-\frac{1}{\sqrt{8v}}-\frac{1}{2}}(2^{3/4}uv^{-3/4})}{D'_{u^2v^{-3/2}2^{-1/2}-\frac{1}{\sqrt{8v}}-\frac{1}{2}}(2^{3/4}uv^{-3/4})}.$$

Nous n'avons pas réussi à trouver ou seulement vérifier que cette expression est solution de (8.3) : y parvenir donnerait une preuve de la proposition 7.3 sans recourir au mouvement brownien du tout.

8.2.4 La formule de Louchard pour les polynômes

Nous allons montrer une nouvelle preuve du théorème de Louchard pour les fonctionnelles polynômiales.

Théorème 8.3 *Soit V une fonction de la forme $V(x)=v_0+v_1|x|+\cdots+v_m|x|^M$ avec $M\in\mathbb{N}$ et $v_0\geqslant 0,v_1\geqslant 0,\cdots,v_M\geqslant 0,|v_0|+|v_1|+\cdots+|v_M|>0$. Soit x le mouvement brownien standard, et soit*

$$g(\alpha):=\int_0^\infty e^{-\alpha t}E_0\left[\exp\left(-\int_0^t V[x(s)]\,ds\right)\middle|x(t)=0\right]\frac{dt}{\sqrt{2\pi t}}.$$

Alors si Y désigne l'excursion brownienne et L sa longueur, on a

$$\int_0^\infty\left(e^{-\alpha t}-1\right)E\left[\exp\left(-\int_0^t V[Y(s)]\,ds\right)\middle|L=t\right]\frac{dt}{\sqrt{2\pi t^3}}=-g^{-1}(\alpha)+\lim_{\alpha\to 0}g^{-1}(\alpha).$$

Corollaire 8.1 *Soit V une fonction de la forme $V(x)=|x|^M$ avec $M\in\mathbb{N}$. Alors si ϕ_B est la transformée de Laplace de la distribution de la fonctionnelle V du pont brownien normalisé, et ϕ_E est la transformée de Laplace de la distribution de la fonctionnelle V de l'excursion normalisée, alors si :*

$$g(\alpha):=\int_0^\infty e^{-\alpha t}\phi_B\left(t^{1+\frac{M}{2}}\right)\frac{dt}{\sqrt{2\pi t}},$$

alors

$$\int_0^\infty\left(e^{-\alpha t}-1\right)\phi_B\left(t^{1+\frac{M}{2}}\right)\frac{dt}{\sqrt{2\pi t^3}}=-g^{-1}(\alpha)+\lim_{\alpha\to 0}g^{-1}(\alpha).$$

Démonstration. Pour un chemin $p=(h_0,h_1,\cdots,h_{2n})$, nous noterons

$$S_M(p)=\sum_{i=0}^{2n}|h_i|^M$$

le paramètre auquel nous nous intéressons. Si $a = (h_0, h_1, \cdots, h_{2n+2})$ est une arche positive, et $d = (h_1, \cdots, h_{2n+1})$ le chemin de Dyck associée, $S_M(a) = \sum_{j=0}^{M} \binom{M}{j} S_j(d)$. Si $D(z, u_1, \cdots, u_M)$ énumère les chemins de Dyck, où chaque u_j correspond au paramètre f_j, $G(z, u, v) = zD(z, u, v)$ vérifie une équation fonctionnelle du type :

$$G(z, u, v) = z + zG(z, u, v)G(zu_1 \cdots u_M, \cdots, u_M),$$

où $G(zu_1 \cdots u_M, \cdots, u_M)/z$ énumère les arches positives — sa $(j+1)^e$ composante est donnée par

$$u_j u_{j+1}^2 \cdots u_{M-1}^{\binom{M-1}{j}} u_M^{\binom{M}{j}} = \prod_{q=j}^{q=M} u_q^{\binom{q}{j}}.$$

En utilisant la même méthodologie que précédemment, nous obtenons par récurrence que pour tout M-uplet $(\alpha_1, \cdots, \alpha_M)$, il existe une constante $A_{\alpha_1, \cdots, \alpha_M}$ telle que, au voisinage de $z = \frac{1}{2}$, si $R = \sqrt{1 - 4z^2}$,

$$z \frac{\partial^{\sum \alpha_i} G(z, u_1, \cdots, u_M)}{\partial u_1^{\alpha_1} \cdots u_M^{\alpha_M}} = \alpha_1! \cdots, \alpha_M! \frac{A_{\alpha_1, \cdots, \alpha_M}}{R^{-1+\sum_{i=1}^{M}(2+M)\alpha_i}} + O\left(\frac{1}{R^{-2+\sum_{i=1}^{M}(2+M)\alpha_i}}\right).$$

Si $\eta_{r_1, \cdots, r_M}$ désigne le moment joint $\mathbf{E}\left[Y_1^{r_1} \cdots Y_M^{r_M}\right]$, où Y_i est la variable aléatoire correspondant à la distribution de la fonctionnelle $x \mapsto |x|^i$ pour l'excursion, une analyse de singularité fournit l'égalité

$$\eta_{r_1, \cdots, r_M} = \frac{4A_{r_1, \cdots, r_M}}{2^{\frac{3r_1 + 4r_2 + \cdots + (2+M)r_M}{2}}} \frac{\sqrt{\pi} r_1! \cdots r_M!}{\Gamma\left(\frac{3r_1 + 4r_2 + \cdots + (2+M)r_M - 1}{2}\right)}.$$

Posons $G_+(z, u_1, \cdots, u_M) = G(zu_1 \cdots u_M, \cdots, u_M)$. On a encore

$$z \frac{\partial^{\sum \alpha_i} G_+(z, u_1, \cdots, u_M)}{\partial u_1^{\alpha_1} \cdots u_M^{\alpha_M}} = \alpha_1! \cdots, \alpha_M! \frac{A_{\alpha_1, \cdots, \alpha_M}}{R^{-1+\sum_{i=1}^{M}(2+M)\alpha_i}} + O\left(\frac{1}{R^{-2+\sum_{i=1}^{M}(2+M)\alpha_i}}\right),$$

et la fonction génératrice P

$$P = \frac{1}{1 - 2zG_+}.$$

On montre par récurrence qu'il existe des constantes $B_{\alpha_1, \alpha_2, \cdots, \alpha_M} > 0$ telles que

$$\frac{\partial^{\sum \alpha_i} P(z, u_1, \cdots, u_M)}{\partial u_1^{\alpha_1} u_2^{\alpha_2} \cdots u_M^{\alpha_M}} \underset{z \to \frac{1}{2}}{\sim} \frac{\alpha_1! \alpha_2! \cdots, \alpha_M! B_{\alpha_1, \alpha_2, \cdots, \alpha_M}}{(\sqrt{1 - 4z^2})^{+1+\sum_{i=1}^{M}(2+i)\alpha_i}}$$

Si $\beta_{r_1, \cdots, r_M}$ désigne le moment joint $\mathbf{E}\left[X_1^{r_1} \cdots X_M^{r_M}\right]$, où X_i est la variable aléatoire correspondant à la distribution de la fonctionnelle $x \mapsto |x|^i$ pour l'excursion, une analyse de singularité fournit l'égalité

$$\beta_{r_1, \cdots, r_M} = \frac{B_{r_1, \cdots, r_M}}{2^{\frac{3r_1 + 4r_2 + \cdots + (2+M)r_M}{2}}} \frac{\sqrt{\pi} r_1! \cdots r_M!}{\Gamma\left(\frac{3r_1 + 4r_2 + \cdots + (2+M)r_M + 1}{2}\right)}.$$

En prenant les termes dominants dans l'égalité

$$(1 - 2zG_+)P = 1$$

dérivée r_i fois par rapport à u_i, on obtient, pour $(r_1, \cdots, r_M) \neq (0, \cdots, 0)$:

$$\sum_{i_1 + j_1 = r_1, \cdots, i_M + j_M = r_M} A_{i_1, \cdots, i_M} B_{j_1, \cdots, j_M} = 0.$$

Définissant les séries

$$A(u_1,\cdots,u_M) := \sum_{r_1,\cdots,r_M} A_{r_1,\cdots,r_M}(-1)^{r_1+\cdots+r_M}u^{r_1}\cdots u^{r_M}$$

et

$$B(u_1,\cdots,u_M) := \sum_{r_1,\cdots,r_M} B_{r_1,\cdots,r_M}(-1)^{r_1+\cdots+r_M}u^{r_1}\cdots u^{r_M},$$

on obtient donc

$$AB(u_1,\cdots,u_M) = \text{ cte } = A(0\cdots,0)B(0\cdots,0) = -\frac{1}{2}.$$

Supposons d'abord $v_0 = 0$. La double transformée de Laplace pour le pont x s'écrit :

$$\begin{aligned}
g(\alpha) &= \int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\int_0^t (V[x(s)]\,ds}\,\middle|\, x(t)=0\right]\frac{dt}{\sqrt{2\pi t}}\\
&= \int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\int_0^t \sum v_k|x(s)|^k\,ds}\,\middle|\, x(t)=0\right]\frac{dt}{\sqrt{2\pi t}}\\
&= \int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\sum v_k t^{1+k/2}\int_0^1 |x(s)|^k\,ds}\,\middle|\, x(1)=0\right]\frac{dt}{\sqrt{2\pi t}}\\
&= \int_0^\infty e^{-\alpha t}\sum\frac{\beta_{r_1,\cdots,r_M}}{r_1!\cdots r_M!}(v_1t^{1+1/2})^{r_1}\cdots(v_Mt^{1+M/2})^{r_M}(-1)^{r_1+\cdots+r_M}\frac{dt}{\sqrt{2\pi t}}\\
&= \sum\frac{\beta_{r_1,\cdots,r_M}}{r_1!\cdots r_M!}(-1)^{r_1+\cdots+r_M}v_1^{r_1}\cdots v_M^{r_M}\int_0^\infty e^{-\alpha t}t^{(1+1/2)r_1+\cdots+(1+M/2)r_M}\frac{dt}{\sqrt{2\pi t}}\\
&= \sum\frac{\beta_{r_1,\cdots,r_M}}{r_1!\cdots r_M!}(-1)^{r_1+\cdots+r_M}v_1^{r_1}\cdots v_M^{r_M}\frac{\Gamma\left((1+1/2)r_1+\cdots+(1+M/2)r_M+1/2\right)}{\sqrt{2\pi}\alpha^{3r_1+4r_2+\cdots+(2+M)r_M+1}}\\
&= \frac{1}{\sqrt{2\alpha}}B\left(\frac{v_1}{(2\alpha)^{1+\frac{1}{2}}},\cdots,\frac{v_k}{(2\alpha)^{1+\frac{k}{2}}},\cdots,\frac{v_M}{(2\alpha)^{1+\frac{M}{2}}}\right)
\end{aligned}$$

Un calcul similaire pour la double transformée de Laplace pour l'excursion Y donne :

$$\begin{aligned}
f'(\alpha) &= -\int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\int_0^t V[Y(s)]\,ds}\,\middle|\, L=t\right]\frac{dt}{\sqrt{2\pi t}}\\
&= \frac{d}{d\alpha}\left(2\sqrt{2\alpha}A\left(\frac{v_1}{(2\alpha)^{1+\frac{1}{2}}},\cdots,\frac{v_k}{(2\alpha)^{1+\frac{k}{2}}},\cdots,\frac{v_M}{(2\alpha)^{1+\frac{M}{2}}}\right)\right)
\end{aligned}$$

Autrement dit, on a $f'(\alpha) = (-1/g(\alpha))'$, ce qui par intégration par rapport à α donne le théorème de Louchard.

Si $v_0 \neq 0$, il suffit de remplacer α par $\alpha + v_0$ dans les calculs ci-dessus.

□

8.3 Arbres non orientés et polyominos parallélogrammes

Rappelons qu'un arbre non orienté (ou non planaire) est un arbre tel qu'à chaque sous-nœud les sous-arbres ne sont pas ordonnés, contrairement aux arbres ordonnés (ou planaires). Quant aux polyominos parallélogrammes, nous renvoyons le lecteur à [20] pour y trouver une définition.

À l'aide des résultats ci-dessus, et du fait que la distribution limite des fonctionnelles aire et moment d'inertie pour les chemins de Dyck est connue, on peut montrer que la distribution limite de ces fonctionnelles est la même pour les arbres non orientés ou pour les polyominos parallélogrammes. La démarche est la suivante : on a trouvé la distribution limite pour des familles simples d'arbres grâce au fait que la distribution du contour de ces arbres converge vers une excursion. On sait qu'on a convergence à la fois en distribution et des moments pour les fonctionnelles qui nous intéressent. La seule connaissance de la convergence des moments nous suffit en fait à affirmer la convergence en distribution, grâce au théorème de Carleman-Lévy : on a une expression explicite de la fonction caractéristique de la distribution limite et on en sait donc suffisamment sur l'asymptotique des moments pour que soit vérifiée la condition de Carleman. Ces moments caractérisent donc uniquement la distribution limite.

En effectuant un pompage des moments sur les séries génératrices des arbres non orientés ou des polyominos (on trouvera ces séries génératrices dans [20, 21, 40, 59]), on retrouve les mêmes types de récurrences sur les termes dominants des séries génératrices des moments de tout ordre. On prouve donc par analyse de singularité la convergence des moments de tout ordre. Il suffit donc de retrouver des récurrences sur les moments identiques à celles qu'on avait trouvées pour les arbres de Catalan ou les chemins de Dyck pour pouvoir affirmer qu'on a la même distribution limite pour l'aire ou le moment d'inertie.

8.3.1 Arbres non orientés

Proposition 8.4 *Soit $V(z,u,v)$ la fonction génératrice des arbres binaires non orientés, z énumérant leur taille, u la longueur de cheminement interne, v la fonctionnelle quadratique. Alors V est donnée par l'équation fonctionnelle*

$$V(z,u,v) = 1 + \frac{1}{2}zV(zuv, uv^2, v)^2 + \frac{1}{2}zV(z^2u^2v^2, u^2v^4, v^2). \tag{8.4}$$

Démonstration. Laissée au lecteur. □

Proposition 8.5 *La fonction génératrice $V(z,1,1)$ énumérant les arbres binaires non orientés admet une singularité dominante ρ de type quadratique, si bien qu'il existe $C > 0$, $\rho > 0$ tels que*

$$[z^n]V(z,1,1) \sim C\frac{\rho^{-n}}{n^{3/2}}.$$

Démonstration. Théorème des fonctions implicites ou l'article d'Otter [175] (qui donne des valeurs approchées de C et ρ). □

Proposition 8.6 *Les moments vérifient la même récurrence que pour les chemins de Dyck.*

Démonstration. Il est clair que $\rho < 1$. Il s'ensuit que la singularité dominante de $V(z^2,1,1)$ est strictement plus grande que ρ, si bien qu'on n'a pas à prendre en compte le terme $V(z^2u^2v^2, u^2v^4, v^2)$ lors du pompage de moments.

Il faut essentiellement remarquer

$$V(z,1,1) = \frac{1 - \sqrt{1 - 4\frac{1}{2}\left(1 + \frac{1}{2}zV(z^2,1,1)\right)}}{z},$$

si bien que

$$1 - zV(z,1,1) = \sqrt{1 - 4\frac{1}{2}\left(1 + \frac{1}{2}zV(z^2,1,1)\right)} = \sqrt{1-\rho z} f(z),$$

avec f analytique au-delà de ρ. Ensuite, le pompage de moments s'effectue de la même manière qu'avant, en dérivant l'égalité 8.4 le nombre idoine de fois par rapport à u et à v, et en faisant $u = v = 1$. □

On peut établir une bijection entre les polyminos parallélogrammes et les chemins de Dyck. Les hauteurs des colonnes dans le polyomino correspondent alors à la hauteur des pics du chemin de Dyck associé [51, 61].

8.3.2 Polyominos

On notera qu'étudier des fonctionnelles telles que la somme des hauteurs ou la somme des carrés des hauteurs des colonnes de polyominos parallélogrammes mènerait à des séries génératrices similaires, aux mêmes récurrences et aux mêmes distributions limites. Nous laissons le lecteur effectuer cette vérification.

Une autre manière de voir cela consiste à remarquer que les pics et vallées dans les chemins de Dyck correspondent respectivement à une montée suivie d'une descente et à une descente suivie d'une montée, et donc à des motifs dans des chemins discrets. En utilisant la même méthodologie que dans [153], on pourrait montrer que le chemin formé des hauteurs de ces motifs converge aussi vers des excursions.

Nous allons maintenant montrer que nous pouvons répondre au passage à la question posée dans [52] qui demandait la nature de la distribution limite du moment d'inertie des polyominos parallélogrammes. Cependant ce que Philippe Duchon appelle moment d'inertie, ou paramètre de rang 3, est un peu différent de ce que nous appelons moment d'inertie. En effet, au lieu de considérer la somme des carrés des hauteurs, Duchon considère la somme des $\binom{k+1}{2}$, où k décrit les hauteurs des colonnes.

Nous allons montrer qu'étudier la somme des $\binom{k+1}{2}$ au lieu des k^2 d'un processus brownien — avant normalisation du processus par $\sqrt{n}$ — revient asymptotiquement au même — comme le conjecturait déjà Philippe Duchon. En effet, si X est le processus de la hauteur d'un processus de branchement ξ de Galton-Watson critique, *i.e.* d'espérance $\mathbf{E}\xi = 1$, et de variance $\sigma^2 = \mathbf{Var}\,\xi \in \mathbb{R}_+^*$, alors

$$\left(\frac{1}{\sqrt{n}} X_n(2nt),\ 0 \leqslant t \leqslant 1\right) \Longrightarrow \left(\frac{2}{\sigma} e(t),\ 0 \leqslant t \leqslant 1\right),$$

et d'après les travaux de Drmota [47], on a

$$\lim_{n\to\infty} \mathbf{E}F\left(\frac{1}{\sqrt{n}} X_n(2nt)\right) = \mathbf{E}F\left(\frac{2}{\sigma} e(t)\right),$$

pour F fonctionnelle polynomialement bornée, si bien que, pour $\beta > 0$, et F intégrale de puissance,

$$\begin{aligned}
\int_0^1 \left(\frac{X_n(2nt)}{\sqrt{n}}\right)^\beta dt &\longrightarrow \left(\frac{2}{\sigma}\right)^\beta \int_0^1 e(t)^\beta\, dt \\
\frac{1}{2n}\int_0^{2n} X_n(s)^\beta\, ds &\longrightarrow \left(\frac{2}{\sigma}\right)^\beta \int_0^1 e(t)^\beta\, dt \\
\frac{1}{2n^{1+\beta/2}} \sum_{i=0}^{2n} X_n(i)^\beta &\longrightarrow \left(\frac{2}{\sigma}\right)^\beta \int_0^1 e(t)^\beta\, dt.
\end{aligned}$$

En particulier, $\sum_{i=0}^{2n} X_n(i)$, qui est de l'ordre de $n^{3/2}$, est négligeable devant $\sum_{i=0}^{2n} X_n(t)^2$, de l'ordre de n^2, et

$$\sum_{i=0}^{2n} \binom{X_n(t)}{2} \underset{n\to\infty}{\sim} \frac{1}{2} \sum_{i=0}^{2n} X_n(i)^2.$$

8.3.3 Des calculs plus simples

Étudier le paramètre de rang 3 au lieu de la somme des carrés des hauteurs simplifie le calcul : pour les chemins de Dyck il mène à des formules du type

$$D(z,u,v) = 1 + z^2 uv D(z,u,v) D(zu, uv, v), \tag{8.5}$$

tandis que le paramètre somme des carrés des hauteurs mène à des formules du type

$$D(z,u,v) = 1 + z^2 uv D(z,u,v) D(zuv, uv^2, v).$$

Le lecteur intéressé pourra donc reprendre les calculs de ce chapitre avec des formules du type (8.5), ou élargir à d'autres familles d'arbres ou classes d'objets combinatoires, avec des équations fonctionnelles plus simples.

8.3.4 Encore d'autres objets ?

Nous avons vu que les mots de Dyck étaient reliés à de nombreux objets combinatoires. On peut se demander si les mots de Dyck bilatères sont reliés à d'autres objets combinatoires, sur lequel l'aire du mot de Dyck bilatère peut se lire de manière sympathique, ou se retrouve, au moins asymptotiquement parlant, dans un paramètre assez naturel de ces autres objets combinatoires. On notera par exemple qu'il existe une bijection entre les mots de Dyck bilatères et les polyominos convexes dirigés [19]. L'aire des mots et le moment d'inertie des mots de Dyck bilatères se retrouvent-ils, au moins asymptotiquement, dans des paramètres assez naturels dans ces autres objets ?

8.4 Méandre et mouvement brownien

Cette section reprend l'article [173].

Nous allons montrer que la manipulation de séries génératrices puis le pompage de moments permet de retrouver la distribution limite de l'aire sous les méandres discrets et la marche de Bernoulli, correspondant à la distribution limite de l'aire sous les méandres (continus) et le mouvement brownien.

Les méandres discrets sont des facteurs gauches de mots de Dyck et peuvent aussi être vus comme des arbres binaires tronqués. De l'étude de l'aire sous les méandres discrets se déduit facilement l'aire sous les marches quelconques, d'où cette digression légèrement hors sujet par rapport à la thèse. En outre, nous pourrons énoncer un théorème similaire au théorème de Louchard, liant cette fois les doubles transformées de la distribution des fonctionnelles polynomiales du mouvement brownien, des méandres et des ponts browniens.

Comme nous l'avons déjà remarqué, dans le monde des fonctions génératrices, l'aire des ponts se déduit de l'aire sous les chemins de Dyck, tandis que l'aire des excursions se déduit de l'aire sous les ponts browniens [140]. Dans le même ordre d'idées, tandis que dans le monde continu l'aire sous le mouvement brownien peut se calculer directement, nous devons d'abord calculer

l'aire sous les chemins de Dyck, les ponts et les méandres discrets avant de trouver l'aire sous les chemins de Bernoulli.

En utilisant la formule de Feynman-Kac, on peut trouver assez facilement la distribution de l'aire signée du mouvement brownien et des ponts browniens [176]. C'est aussi le cas en utilisant la "generatingfunctionology" (la technologie des fonctions génératrices).

8.4.1 Fonctions génératrices

Soient D, F, P et B les fonctions génératrices de respectivement les chemins de Dyck, les méandres, les chemins de Grand Dyck, et les marches aléatoires de Bernoulli (avec des pas -1 et -1), où z énumère la longueur des marches, et u son aire. Nous appellerons chemin de Dyck *positif* un chemin de Dyck classique, et chemin de Dyck *négatif* un chemin tel que son symétrique par rapport à l'axe des x soit un chemin de Dyck positif. De la même manière, nous définissons les notions de méandres positifs et négatifs. Nous avons

Proposition 8.7

$$D(z,u) = 1 + z^2 u D(z,u) D(zu,u) \tag{8.6}$$
$$F(z,u) = D(z,u) + uzD(z,u)F(zu,u) \tag{8.7}$$
$$P(z,u) = \frac{1}{1-2uz^2 D(zu,u)} \tag{8.8}$$
$$B(z,u) = P(z,u)(2zuF(zu,u)+1) \tag{8.9}$$

Démonstration. Un chemin de Dyck est soit un chemin de Dyck vide, énuméré par $1 \cdot z^0 = 1$, soit un chemin de Dyck, énuméré par $D(z,u)$, suivi d'un chemin de Dyck non vide, énuméré par $z^2 u D(zu,u)$.

Un méandre est soit un chemin de Dyck, *i.e.* un méandre se terminant en zéro, énuméré par $D(z,u)$, soit un chemin de Dyck, énuméré par $D(z,u)$, suivi d'un méandre ne se terminant pas en zéro, énuméré par $uzF(zu,u)$.

Un chemin de Grand Dyck est une suite de chemins de Dyck non vides positifs ou négatifs, énumérés par $2uz^2 D(zu,u)$.

Un chemin de Bernoulli soit se termine en zéro, auquel cas c'est un chemin de Grand Dyck, énuméré par $P(z,u)$, soit ne se termine pas en zéro, auquel cas c'est un chemin de Grand Dyck, énuméré par $P(z,u)$, suivi d'un méandre strictement positif ou négatif, énuméré par $2zuF(zu,u)$. □

Remarque. Pour les méandres ou les chemins de Bernoulli, le paramètre "somme des valeurs absolues des hauteurs" peut différer de l'aire comprise entre l'axe des abscisses et le chemin, d'un nombre de l'ordre de la racine carrée du nombre de retours en zéro. Néanmoins, ce paramètre et la véritable aire possèdent asymptotiquement la même distribution.

8.4.2 Chemins de Dyck et chemins de Grand Dyck

Nous rappelons ici les résultats concernant la distribution limite de l'aire de chemins de Dyck et des chemins de Grand Dyck. Les schémas de preuve fourniront des notations utiles pour la section suivante.

Le lecteur intéressé par des preuves plus complètes pourra soit reprendre la méthodologie utilisée en début de chapitre dans le cas multivarié, soit lire en-dessous, pour le cas multivarié, les preuves des lemme 8.1 et théorème 8.3.

Si B_s est un mouvement brownien, alors la propriété de scaling $B_s \overset{\mathcal{D}}{\equiv} \sqrt{\ell} B_{s/\ell}$, pour tout ℓ, aussi valide pour les méandres, ponts et excursions (ℓ est alors dans ces deux derniers cas une variable aléatoire égale à la longueur du mouvement), implique

$$\int_0^t |B_s|\, ds \overset{\mathcal{D}}{\equiv} t^{3/2} \int_0^1 |B_s|\, ds, \tag{8.10}$$

et des formules similaires pour les méandres, ponts et excursions, conditionnées à avoir longueur 1 dans le membre de droite. Donc, en ce qui concerne les chemins de Dyck et les chemins de Grand Dyck de longueur $2n$, il est naturel de normaliser leur aire par $(2n)^{3/2}$. Il serait facile de vérifier, à l'aide des calculs ci-après pour les moments de tout ordre entier, que la variance de l'aire des divers chemins considérés est en effet d'ordre $\sigma^2 n^3$ avec $\sigma^2 > 0$, quoique $\sigma \neq 2$.

Théorème 8.1 *Soit ϕ_D la transformée de Laplace de l'aire des excursions conditionnés à avoir longueur 1. Alors elle vérifie :*

$$\int_0^\infty (e^{-us} - 1)\phi_D(\sqrt{2}s^{3/2})\, \frac{ds}{s^{3/2}} = 2\sqrt{\pi}\left(\frac{Ai'(u)}{Ai(u)} - \frac{Ai'(0)}{Ai(0)}\right). \tag{8.11}$$

Démonstration. Nous rappelons dans les grandes lignes la preuve de Takács [200]. En dérivant (8.6) r fois par rapport à u et en faisant $u = 1$, on peut montrer par récurrence qu'il existe pour tout r une constante K_r non nulle telle que

$$\frac{D_{u^r}}{r!} \underset{z^2 \to \frac{1}{4}}{\sim} \frac{2^{r+2} K_r}{(1 - 4z^2)^{\frac{3r-1}{2}}}. \tag{8.12}$$

En outre, la suite (K_r) vérifie la récurrence

$$K_r = \sum_{i=1}^{r-1} K_i K_{r-i} - K_0 K_{r-1}(3r - 4)/2 \tag{8.13}$$

avec $K_0 = -1/2$. La série $y(z) = \sum K_r (-z)^r$ est alors solution de l'équation

$$3z^2 y'(z) - zy(z) = 4y^2(z) - 1 \tag{8.14}$$

qui (étant donné $y(0) = -1/2$) se résout en

$$y(z) = Ai'(z^{-2/3}) z^{1/3} \Big/ [2Ai(z^{-2/3})], \tag{8.15}$$

où Ai est la fonction d'Airy, qui est la solution de $w'' = xw$ telle que $\lim_{x\to\infty} w(x) = 0$ (p. 446 de [1]). Si M_n^D est le moment d'ordre n de la distribution limite de l'aire des chemins de Dyck normalisé par $(\sqrt{2n})^3$, alors par analyse de singularité [81]

$$M_r^D = \lim_{n\to\infty} \frac{[z^{2n}] D_{u^r}(z, 1)}{(\sqrt{2n})^{3r} [z^{2n}] D(z, 1)} = \lim_{n\to\infty} \frac{[z^{2n}] \frac{2^{r+2} K_r}{(1-4z^2)^{\frac{3r-1}{2}}}}{(\sqrt{2n})^{3r} [z^{2n}](-2\sqrt{1 - 4z^2})} = \frac{4\sqrt{\pi} K_r r!}{\Gamma\left(\frac{3r-1}{2}\right) 2^{\frac{r}{2}}}. \tag{8.16}$$

En utilisant

$$\phi_D(x) = \sum \frac{M_r^D}{r!} (-x)^r, \tag{8.17}$$

il est facile de trouver (8.11) en faisant une intégration par parties. □

La distribution de l'aire sous les excursions étant maintenant connue, nous allons maintenant déterminer la distribution de l'aire sous les ponts.

Théorème 8.2 *Soit ϕ_P la transformée de Laplace de la distribution de la valeur absolue de l'aire des ponts browniens conditionnés à avoir longueur 1. Alors elle vérifie*

$$\int_0^\infty e^{-us}\phi_P(\sqrt{2}s^{3/2})s^{-1/2}\,ds = -\sqrt{\pi}\frac{Ai(u)}{Ai'(u)}. \tag{8.18}$$

Démonstration. Nous procédons comme dans [174]. En différentiant

$$\left(1-2uz^2D(zu,u)\right)P(z,u)=1 \tag{8.19}$$

r fois par rapport à u et en faisant $u=1$, on peut montrer qu'il existe une suite (Q_r) de nombres positifs telle que

$$\frac{1}{r!}P_{u^r}(z,1) \underset{z^2\to\frac{1}{4}}{\sim} \frac{2^rQ_r}{(\sqrt{1-4z^2})^{3r+1}}. \tag{8.20}$$

Elle est donnée par

$$Q_0=1 \quad \text{and} \quad \forall r\geqslant 1,\ Q_r=2\sum_{j=1}^{r}K_jQ_{r-j}. \tag{8.21}$$

La fonction $q(z)=\sum_{r\geqslant 0}Q_r(-z)^r$ vérifie

$$q(z)=-\frac{1}{2y(z)}=-\frac{Ai(z^{-2/3})}{Ai'(z^{-2/3})z^{1/3}}. \tag{8.22}$$

Définissant (M_r^P) telle que $\phi_P(x)=\sum\frac{M_r^P}{r!}(-x)^r$, on obtient par analyse de singularité

$$M_r^P=\lim_{n\to\infty}\frac{[z^{2n}]\,P_{u^r}(z,1)}{(2n)^{\frac{3r}{2}}\,[z^{2n}]\,P(z,1)}=\sqrt{\pi}\frac{Q_rr!}{2^{r/2}\Gamma\left(\frac{3r+1}{2}\right)}, \tag{8.23}$$

et il est alors facile d'établir la formule (8.18). □

8.4.3 Méandre et marche de Bernoulli

Dans cette section, nous retrouvons la distribution de l'aire des méandres et des marches de Bernoulli. Comme cela n'a pas été fait auparavant avec le seul usage de fonctions génératrices, nous allons effectuer les calculs de manière plus détaillée que dans la section précédente.

Lemme 8.1 *Il existe pour tout r des constantes non nulles a_r^D, a_r^F, a_r^P, et a_r^B, telles que*

$$\frac{D_{u^r}}{r!} \underset{z\to\frac{1}{2}}{\sim} \frac{a_r^D}{(1-2z)^{\frac{3r-1}{2}}} \tag{8.24}$$

$$\frac{F_{u^r}}{r!} \underset{z\to\frac{1}{2}}{\sim} \frac{a_r^F}{(1-2z)^{\frac{3r+1}{2}}} \tag{8.25}$$

$$\frac{P_{u^r}}{r!} \underset{z\to\frac{1}{2}}{\sim} \frac{a_r^P}{(1-2z)^{\frac{3r+1}{2}}} \tag{8.26}$$

$$\frac{B_{u^r}}{r!} \underset{z\to\frac{1}{2}}{\sim} \frac{a_r^B}{(1-2z)^{\frac{3r+2}{2}}} \tag{8.27}$$

où $\left(a_n^F\right)$ *est donnée par la condition initiale* $a_0^F = \sqrt{2}$ *et la récurrence*

$$\forall r \geqslant 1,\ \sqrt{2}a_r^F = \frac{1}{4}(3r-2)a_{r-1}^F + \frac{1}{2}\sum_{j=0}^{r-1} a_{r-j}^D a_j^F. \tag{8.28}$$

En outre, $\left(a_n^B\right)$ *vérifie*

$$\forall r \geqslant 0,\ a_r^B = \sum_{j=0}^{r} a_{r-j}^P a_j^F. \tag{8.29}$$

Démonstration. D'après les résultats précédents, a_r^D et a_r^P existent de manière évidente et sont égaux respectivement à $2^{r+2}K_r$ et $2^{-\frac{r+1}{2}}Q_r$, ce qui prouve (8.24) et (8.26).

Montrons (8.25). Nous avons

$$F(z,1) = \frac{-1+\frac{1+2z}{1-2z}}{2z} \underset{z\to\frac{1}{2}}{\sim} \frac{\sqrt{2}}{\sqrt{1-2z}}, \tag{8.30}$$

si bien que (8.25) est vrai pour $r = 0$.

Si on différencie r fois l'égalité (8.7), on obtient

$$F_{u^r} = D_{u^r} + z\sum_{j=0}^{r}\binom{r}{j} D_{u^{r-j}}\left(uF(zu,u)\right)^{(j)}, \tag{8.31}$$

avec

$$\left(uF(zu,u)\right)^{(j)} = u(F_{u^j} + jzF_{zu^{j-1}} + \cdots) + 1\cdot(\cdots). \tag{8.32}$$

Donc en $u = 1$

$$(1-zD)F_{u^r} = rz^2 DF_{zu^{r-1}} + z\sum_{j=0}^{r-1}\binom{r}{j} D_{u^{r-j}}F_{u^j} + \cdots, \tag{8.33}$$

où les termes mis en points de suspension sont négligeables par rapport à ceux qui sont écrits lorsque z tend vers 2.

Soit $R = (\sqrt{1-2z})^{-1}$. Les D_{u^j} et F_{u^j} ont des termes dominants en R respectivement de degrés $3(r-j)-1$ et $3j+1$, si bien que $D_{u^j}F_{u^{r-j}}$ est de degré $3r$. En outre, l'asymptotique suivante, pour z tendant vers $\frac{1}{2}$

$$\frac{F_{u^{r-1}}}{(r-1)!} \sim \frac{a_{r-1}^F}{(1-2z)^{\frac{3(r-1)+1}{2}}} = \frac{a_{r-1}^F}{(1-2z)^{\frac{3r-2}{2}}} \tag{8.34}$$

entraîne

$$\frac{F_{zu^{r-1}}}{(r-1)!} \sim \frac{(3r-2)a_{r-1}^F}{(1-2z)^{\frac{3r}{2}}} \tag{8.35}$$

si bien que $F_{zu^{r-1}}$ est de degré $3r$. Comme nous avons

$$1 - zD \underset{z\to\frac{1}{2}}{\sim} \sqrt{2}\sqrt{1-2z}, \tag{8.36}$$

on en déduit que $(1 - zD)F_{u^r}$ est de degré $3r$, si bien que F_{u^r} est de degré $3r + 1$ et (8.25) est prouvé.

Nous avons $(z^2 D)(2) = \frac{1}{2}$ donc, en identifiant les coefficients du monôme R^{3n}, nous obtenons

$$\sqrt{2} a_r^F r! = \frac{1}{2} r(r-1)!(3r-2) a_{r-1}^F + \frac{1}{2} \sum_{j=0}^{r-1} \binom{r}{j} (r-j)! a_{r-j}^D j! a_j^F , \tag{8.37}$$

qui, divisé par $r!$, devient (8.28).

Nous montrons (8.27) et (8.29) de la même manière. En dérivant (8.9) r fois, nous obtenons

$$B_{u^r}(z,1) = \sum_{j=0}^{r} \binom{r}{j} P_{u^{r-j}}(z,1) F_{u^j}(z,1) + \cdots , \tag{8.38}$$

ce qui nous donne (8.27) avec (a_r^B) vérifiant (8.29). □

Nous pouvons maintenant déterminer les doubles transformées de Laplace de la distribution de l'aire des méandres et du mouvement brownien. Remarquons que la longueur n des méandres discrets et des marches de Bernoulli peut être aussi bien paire qu'impaire, si bien que nous normalisons la variable aléatoire par $n^{3r/2}$.

Théorème 8.3 *Soit ϕ_F la transformée de Laplace de la distribution de l'aire absolue des méandres browniens de longueur 1. Elle vérifie*

$$\int_0^\infty e^{-us} \phi_F\left(\sqrt{2s^3}\right) s^{-1/2}\, ds = -\sqrt{\pi} \frac{3 \int_0^u Ai(t)\, dt - 1}{3 Ai(u)}. \tag{8.39}$$

Démonstration. Comme $D \sim -\sqrt{2} R$, nous avons $a_0^D = -\sqrt{2}$, si bien que (8.28) peut se réécrire

$$\frac{3}{2}(r-1) a_{r-1}^F + \frac{1}{2} a_{r-1}^F + \frac{1}{2} \sum_{j=0}^{r} a_{r-j}^D a_j^F = 0. \tag{8.40}$$

Si nous multiplions chaque égalité par $(-x)^r$ et sommons les égalités obtenues de $r = 1$ à l'infini, nous obtenons

$$\frac{3}{2} x^2 A_F' - 2x A_F + \frac{1}{2}(A_F A_D - a_0^F a_0^D) = 0, \tag{8.41}$$

où

$$A_F(x) = \sum_{j=0}^{r} a_r^F (-x)^r \quad \text{and} \quad A_D(x) = \sum_{j=0}^{r} a_r^D (-x)^r. \tag{8.42}$$

Comme nous avons $a_0^F = \sqrt{2}$, (8.41) peut se réécrire

$$-\frac{3}{2} x^2 A_F' + \left(\frac{A_D}{2} - \frac{1}{2} x\right) A_F + 2 = 0, \tag{8.43}$$

et, en remplaçant x par $\frac{1}{2z^{3/2}}$, nous obtenons

$$-\frac{3}{2} \frac{1}{4z^3} A_F'\left(\frac{1}{2z^{3/2}}\right) + \left(\frac{1}{2} A_D\left(\frac{1}{2z^{3/2}}\right) - \frac{1}{4z^{3/2}}\right) A_F\left(\frac{1}{2z^{3/2}}\right) + 2 = 0. \tag{8.44}$$

Soit $f(z) = A_F\left(\frac{\sqrt{2}}{z^{3/2}}\right)$. Cette fonction vérifie

$$f'(z) = -\frac{3\sqrt{2}}{z^{3/2}} A'_F\left(\frac{\sqrt{2}}{z^{3/2}}\right). \tag{8.45}$$

Nous savons déjà

$$A_D(x) = 4\sqrt{2}y\left(\frac{x}{\sqrt{2}}\right) = 4\sqrt{2}\frac{Ai'\left(\left(\frac{x}{\sqrt{2}}\right)^{-2/3}\right)}{2Ai\left(\left(\frac{x}{\sqrt{2}}\right)^{-2/3}\right)}, \tag{8.46}$$

si bien que

$$A_D\left(\frac{\sqrt{2}}{z^{3/2}}\right) = 2\sqrt{2}\frac{Ai'(z)z^{-1/2}}{Ai(z)}, \tag{8.47}$$

et

$$f'(z) = \left(\frac{1}{2z} - \frac{Ai'(z)}{Ai(z)}\right) f(z) - \sqrt{2z}. \tag{8.48}$$

Résolvons cette équation, avec la condition initiale $f(\infty) = \sqrt{2}$, à l'aide de méthode de la variation de la constante. L'équation homogène associée s'écrit

$$\frac{f'}{f} = \frac{1}{2z} - \frac{Ai'(z)}{Ai(z)}$$

d'où

$$\log f = \frac{1}{2}\log z - \log Ai = \log \frac{z^{1/2}}{Ai}$$

et

$$f = \lambda \frac{z^{1/2}}{Ai}.$$

Ensuite,

$$\lambda' \frac{z^{1/2}}{Ai} = -\sqrt{2z}$$

entraîne

$$\lambda(z) = -\sqrt{2}\left(\int_0^z Ai(u)\ du - C\right),$$

donc

$$f(z) = -\sqrt{2z}\frac{\int_0^z Ai(u)\ du - C}{Ai(z)}.$$

et la condition $f(\infty) = \sqrt{2}$ entraîne $C = \frac{1}{3}$, étant donné la condition initiale $f(\infty) = \sqrt{2}$ et les asymptotiques **10.4.59** et **10.4.82** de [1], pour x grand

$$Ai(x) \sim \frac{1}{2}\pi^{-1/2}x^{-1/4}e^{-\frac{2}{3}x^{3/2}}, \tag{8.49}$$

$$\int_0^x Ai(t)\ dt \sim \frac{1}{3} - \frac{1}{2}\pi^{-1/2}x^{-3/4}\exp\left(-\frac{2}{3}x^{3/2}\right). \tag{8.50}$$

On trouve donc finalement

$$f(z) = -\sqrt{2z}\frac{3\int_0^z Ai(u)\ du - 1}{3Ai(z)}. \tag{8.51}$$

En effectuant une analyse de singularité sur $F_{u^r}(z,1)$, nous trouvons

$$M_r^F = \lim_{n\to\infty} \frac{[z^n]\, F_{u^r}(z,1)}{n^{\frac{3r}{2}}\,[z^n]\, F(z,1)} = \frac{\sqrt{\pi}r!a_r^F}{\sqrt{2}\Gamma\left(\frac{3r+1}{2}\right)}. \tag{8.52}$$

Finalement

$$\begin{aligned}
\int_0^\infty e^{-us}\phi_F\left(\sqrt{2s^3}\right)s^{-1/2}\ ds &= \int_0^\infty \sum_{r\geqslant 0}\frac{M_r^F}{r!}(-\sqrt{2}s^{3/2})^r s^{-1/2}e^{-us}\ ds && (8.53)\\
&= \frac{\pi}{2}\sum_{r\geqslant 0}\int_0^\infty \frac{a_r^F}{\Gamma\left(\frac{3r+1}{2}\right)}(-\sqrt{2}s^{3/2})^r s^{-1/2}e^{-us}\ ds && (8.54)\\
&= \frac{\pi}{2}\sum_{r\geqslant 0} a_r^F \frac{(\sqrt{2})^r}{u^{(3r+1)/2}} = \sqrt{\frac{\pi}{2u}}A_F\left(\frac{\sqrt{2}}{u^{3/2}}\right) && (8.55)\\
&= -\sqrt{\pi}\frac{3\int_0^u Ai(t)\ dt - 1}{3Ai(u)}. \qquad \square && (8.56)
\end{aligned}$$

Théorème 8.4 *Soit ϕ_B la transformée de Laplace de la distribution de l'aire absolue du mouvement brownien de longueur 1. Elle vérifie*

$$\int_0^\infty e^{-zt}\phi_B\left(\sqrt{2t^3}\right)\ dt = \frac{3\int_0^z Ai(u)\ du - 1}{3Ai'(z)}. \tag{8.57}$$

Démonstration. Soit $A_P(x) = \sum_{n\geqslant 0} a_n^P(-x)^n$, qui est donné par

$$A_P(x) = \frac{1}{\sqrt{2}}g\left(\frac{x}{\sqrt{2}}\right) = -\frac{1}{\sqrt{2}}\frac{2Ai\left(\left(\frac{x}{\sqrt{2}}\right)^{-2/3}\right)}{Ai'\left(\left(\frac{x}{\sqrt{2}}\right)^{-2/3}\right)} = -\frac{1}{\sqrt{2}}\frac{Ai(z)z^{1/2}}{Ai'(z)}. \tag{8.58}$$

De (8.29) nous déduisons

$$A_B(x) = A_P(x)A_B(x), \tag{8.59}$$

si bien que

$$\begin{aligned}
A_B\left(\frac{\sqrt{2}}{z^{3/2}}\right) &= A_P\left(\frac{\sqrt{2}}{z^{3/2}}\right)A_F\left(\frac{\sqrt{2}}{z^{3/2}}\right) && (8.60)\\
&= \left(-\frac{1}{\sqrt{2}}\frac{Ai(z)z^{1/2}}{Ai'(z)}\right)\cdot\left(\sqrt{2z}\frac{3\int_0^z Ai(u)\ du - 1}{3Ai(z)}\right) && (8.61)\\
&= z\frac{3\int_0^z Ai(u)\ du - 1}{3Ai'(z)}, && (8.62)
\end{aligned}$$

d'où

$$\begin{aligned}
\int_0^\infty e^{-zt}\Phi_B\left(\sqrt{2t^3}\right)\,dt &= \int_0^\infty e^{-zt}\sum_{n\geqslant 0}\frac{M_{Bn}}{n!}(-1)^n(2t^3)^{\frac{n}{2}}\,dt \\
&= \sum_{n\geqslant 0}\frac{d_{Bn}}{\Gamma\left(\frac{3n+2}{2}\right)}(-1)^n\int_0^\infty e^{-zt}(2t^3)^{\frac{n}{2}}\,dt \\
&= \sum_{n\geqslant 0}\frac{d_{Bn}}{\Gamma\left(\frac{3n+2}{2}\right)}2^{\frac{n}{2}}z^{-\frac{3n}{2}-1}\Gamma\left(\frac{3n+2}{2}\right)\,dt \\
&= \sum_{n\geqslant 0}d_{Bn}\left(-\frac{\sqrt{2}}{z^{3/2}}\right)^n\frac{1}{z} \\
&= \frac{1}{z}A_B\left(\frac{\sqrt{2}}{z^{3/2}}\right) \\
&= \frac{3\int_0^z Ai(u)\,du-1}{3Ai'(z)}.
\end{aligned}$$

□

8.4.4 Aires signées

Nous avons jusqu'à maintenant uniquement considéré des fonctionnelles ne dépendant que de la distance à l'axe des abscisses de la particule décrivant un mouvement brownien. Or on peut très bien prendre en compte le fait que la particule se déplace au-dessus ou en-dessous de l'axe des x. Dans [176] se trouve le théorème suivant, qui donne la distribution jointe de l'aire du mouvement brownien respectivement au-dessus et en-dessous de l'axe des x, en utilisant le théorème de Feynman-Kac.

Théorème 8.5 *Définissons*

$$x^+ = \begin{cases} x & si\ x\geqslant 0, \\ 0 & sinon, \end{cases} \quad et \quad x^- = \begin{cases} -x & si\ x\leqslant 0, \\ 0 & sinon. \end{cases} \tag{8.63}$$

Soit V la fonction $x\mapsto \beta x^+ + \gamma x^-$. Alors pour $\alpha,\beta,\gamma>0$ nous avons

$$\begin{aligned}
&\int_0^\infty e^{-\alpha t}\mathbf{E}\left[e^{-t^{3/2}\int_0^1 V(b_s)\,ds}\right]\frac{dt}{\sqrt{2\pi t}} \\
&= -\frac{2Ai((2\gamma)^{1/3}\alpha/\gamma)Ai((2\beta)^{1/3}\gamma/\alpha)}{(2\gamma)^{1/3}Ai'((2\gamma)^{1/3}\alpha/\gamma)Ai((2\beta)^{1/3}\alpha/\beta)+(2\beta)^{1/3}Ai'((2\beta)^{1/3}\alpha/\beta)Ai((2\gamma)^{1/3}\alpha/\gamma)}
\end{aligned} \tag{8.64}$$

et, si W_t désigne le mouvement brownien standard,

$$\begin{aligned}
&\int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\int_0^t V(W_s)\,ds}\right]\,dt \\
&= -\frac{2Ai((2\beta)^{1/3}\alpha/\beta)\int_0^\infty Ai((2\gamma)^{1/3}(x+\alpha/\gamma))\,dx}{(2\gamma)^{1/3}Ai((2\gamma)^{1/3}\alpha/\gamma)Ai((2\beta)^{1/3}\alpha/\beta)+(2\beta)^{1/3}Ai((2\beta)^{1/3}\alpha/\beta)Ai((2\gamma)^{1/3}\alpha/\gamma)} \\
&\quad -\frac{2Ai((2\gamma)^{1/3}\alpha/\gamma)\int_0^\infty Ai((2\beta)^{1/3}(x+\alpha/\beta))\,dx}{(2\gamma)^{1/3}Ai((2\gamma)^{1/3}\alpha/\gamma)Ai((2\beta)^{1/3}\alpha/\beta)+(2\beta)^{1/3}Ai((2\beta)^{1/3}\alpha/\beta)Ai((2\gamma)^{1/3}\alpha/\gamma)}.
\end{aligned} \tag{8.65}$$

En utilisant le cadre précédent, il est possible de donner une autre preuve de ce théorème.
Résumé de preuve du théorème 8.5. Si u_+ énumère l'aire d'un chemin au-dessus de l'axe des x, et u_- l'aire d'un chemin en-dessous de l'axe des x, alors les fonctions génératrices P et B des chemins de Grand Dyck et de Bernoulli vérifient

$$P(z,u_+,u_-) = \frac{1}{1-z^2\left(u_+D(zu_+,u_+)+u_-D(zu_-,u_-)\right)} \tag{8.66}$$

$$B(z,u_+,u_-) = P(z,u_+,u_-)\left(1+z\sqrt{u_+}F(zu_+,u_+)+z\sqrt{u_-}F(zu_-,u_-)\right) \tag{8.67}$$

Si nous définissons Q_{rs} de telle sorte que

$$\frac{P_{u_+^r u_-^s}(z,1)}{r!s!} \underset{z\to\frac{1}{2}}{\sim} \frac{2^{r+s}Q_{rs}}{(1-2z)^{\frac{3r+1}{2}}}, \tag{8.68}$$

et $q(u,v) = \sum Q_{rs}(-u)^r(-v)^s$, alors après dérivation

$$\left(1-z^2\left(u_+D(zu_+,u_+)+u_-D(zu_-,u_-)\right)\right)P(z,u_+,u_-) = 1 \tag{8.69}$$

par rapport à u_+ et u_-, et identification des coefficients dominants en $(1-4z^2)^{-1/2}$, nous obtenons

$$q(u_-,u_+) = -\frac{1}{y(u_-)+y(u_+)}, \tag{8.70}$$

qui se réécrit en (8.64) avec $u_+ = \beta\cdot 2^{-1/2}\alpha^{-3/2}$ et $u_- = \gamma\cdot 2^{-1/2}\alpha^{-3/2}$.

(8.65) peut se prouver de la même manière, en partant de (8.67). □

Notons que la relation (8.70) se retrouve aussi dans l'exercice 4.18 du chapitre XIII de [183], pour toute fonctionnelle additive. Par pompage de moments, nous pourrions aussi étendre cette relation à des fonctionnelles de type polynomial, en effectuant des calculs similaires à la section suivante.

Mentionnons également la référence [203] qui traite de l'asymptotique de la partie positive d'un pont brownien.

8.4.5 Autres fonctionnelles

Dans [174], le théorème de Louchard est prouvé pour des fonctionnelles polynomiales, en utilisant une version multivariée de (8.8). Il est possible d'utiliser la même méthode avec (8.9) pour établir un théorème similaire, liant le mouvement brownien standard, les méandres et les ponts browniens.

Théorème 8.6 *Notons m le méandre et, comme avant, b le pont brownien et W le mouvement brownien standard. Pour des réels strictement positifs $h_0, h_1, \cdots, h_M$ tels que $|h_1|+\cdots+|h_M| \neq 0$, posons $h : x \mapsto \sum_{j=0}^M h_j|x|^j$. Soient g_s, g_b et g_m des doubles transformées de Laplace de la distribution de la fonctionnelle h pour respectivement le mouvement brownien standard partant de 0, le pont brownien et le méandre :*

$$g_s(\alpha) = \int_0^\infty e^{-\alpha t}\mathbf{E}_0\left[e^{-\int_0^t h(W_s)\,ds}\right]\,dt \tag{8.71}$$

$$g_b(\alpha) = \int_0^\infty e^{-\alpha t}\mathbf{E}\left[e^{-\int_0^t h(b_s)\,ds}\right]\frac{dt}{\sqrt{2\pi t}} \tag{8.72}$$

$$g_m(\alpha) = \int_0^\infty e^{-\alpha t}\mathbf{E}\left[e^{-\int_0^t h(m_s)\,ds}\right]\frac{dt}{\sqrt{2\pi t}} \tag{8.73}$$

Alors

$$g_s = 2g_b g_m. \tag{8.74}$$

Démonstration. Soit $M \in \mathbb{N}^*$. Notons u_k la variable énumérant la somme des puissances $k$èmes de la valeur absolue des altitudes de la chemin. Alors

$$B(z, u_1, \cdots, u_M) = P(z, u_1, \cdots, u_M)(2zu_1 \cdots u_M)F(zu_1, \cdots, u_M) + 1), \tag{8.75}$$

où la $(j+1)$ème composante de F est donnée par

$$u_j u_{j+1}^2 \cdots u_{M-1}^{\binom{M-1}{j}} u_M^{\binom{M}{j}} = \prod_{q=j}^{q=M} u_q^{\binom{q}{j}}. \tag{8.76}$$

Soit $a^B_{r_1,r_2,\cdots,r_M}$, $a^P_{r_1,r_2,\cdots,r_M}$, $a^F_{r_1,r_2,\cdots,r_M}$ les constantes telles que, pour tout $(r_1, r_2, \cdots, r_M) \in \mathbb{N}^M$,

$$\frac{F_{u_1^{r_1}u_2^{r_2}\cdots u_M^{r_M}}(z,1)}{r_1!r_2!\cdots r_M!} \underset{z\to\frac{1}{2}}{\sim} \frac{a^F_{r_1,r_2,\cdots,r_M}}{(1-2z)^{\frac{1+\sum(2+i)r_i}{2}}} \tag{8.77}$$

$$\frac{P_{u_1^{r_1}u_2^{r_2}\cdots u_M^{r_M}}(z,1)}{r_1!r_2!\cdots r_M!} \underset{z\to\frac{1}{2}}{\sim} \frac{a^P_{r_1,r_2,\cdots,r_M}}{(1-2z)^{\frac{1+\sum(2+i)r_i}{2}}} \tag{8.78}$$

$$\frac{B_{u_1^{r_1}u_2^{r_2}\cdots u_M^{r_M}}(z,1)}{r_1!r_2!\cdots r_M!} \underset{z\to\frac{1}{2}}{\sim} \frac{a^B_{r_1,r_2,\cdots,r_M}}{(1-2z)^{\frac{2+\sum(2+i)r_i}{2}}}. \tag{8.79}$$

Soit

$$A_B(u_1, \cdots, u_M) = \sum a^B_{r_1,r_2,\cdots,r_M}(-u_1)^{r_1}\cdots(-u_M)^{r_M}, \tag{8.80}$$

$$A_P(u_1, \cdots, u_M) = \sum a^P_{r_1,r_2,\cdots,r_M}(-u_1)^{r_1}\cdots(-u_M)^{r_M}, \tag{8.81}$$

$$A_F(u_1, \cdots, u_M) = \sum a^F_{r_1,r_2,\cdots,r_M}(-u_1)^{r_1}\cdots(-u_M)^{r_M}. \tag{8.82}$$

Alors (8.75) se réécrit $A_B = A_P A_F$. Comme nous avons

$$g_s(\alpha) = \frac{1}{\alpha} A_B\left(\frac{h_1}{\alpha^{1+\frac{1}{2}}}, \cdots, \frac{h_k}{\alpha^{1+\frac{k}{2}}}, \cdots, \frac{h_M}{\alpha^{1+\frac{M}{2}}}\right) \tag{8.83}$$

$$g_b(\alpha) = \frac{1}{\sqrt{\alpha}} A_P\left(\frac{h_1}{\alpha^{1+\frac{1}{2}}}, \cdots, \frac{h_M}{\alpha^{1+\frac{M}{2}}}\right), \tag{8.84}$$

$$g_m(\alpha) = \frac{1}{2\sqrt{\alpha}} A_F\left(\frac{h_1}{\alpha^{1+\frac{1}{2}}}, \cdots, \frac{h_M}{\alpha^{1+\frac{M}{2}}}\right), \tag{8.85}$$

cela établit le théorème pour $h_0 = 0$. Pour traiter le cas $h_0 \neq 0$, il suffit de remplacer α pour $\alpha + h_0$. □

Ce résultat reste vrai pour toute fonctionnelle paire borélienne, ce qu'on peut montrer en utilisant le co-méandre brownien [183, 214]

$$\tilde{m}_u = \frac{1}{1-g_1}\left|B_{g_1+u(1-g_1)}\right|,$$

indépendant de $\mathcal{F}_{g_1}$, et en utilisant, dans le monde continu, une décomposition similaire à la décomposition combinatoire utilisée ici. Un mouvement brownien de longueur d_1 (temps de

premier retour à zéro après l'instant 1) est un pont de longueur g_1 (temps de dernier passage à zéro avant l'instant 1), suivi du co-méandre défini ci-dessus.

Pour la fonctionnelle $h(x) = x^2$, en injectant les résultats connus (*via* les objets continus)

$$\mathbf{E}_0\left[e^{-\int_0^t W_s^2\, ds}\right] = \left(\cosh(t\sqrt{2})\right)^{-1/2}, \tag{8.86}$$

$$\mathbf{E}\left[e^{-\int_0^t b_s^2\, ds}\right] = \left(t\sqrt{2}/\sinh(t\sqrt{2})\right)^{1/2}, \tag{8.87}$$

$$\mathbf{E}\left[e^{-\int_0^t m_s^2\, ds}\right] = \left(\cosh(t\sqrt{2})\right)^{-1}\left(t\sqrt{2}/\sinh(t\sqrt{2})\right)^{1/2}, \tag{8.88}$$

dans (8.74) et en inversant en α les transformées de Laplace de chaque côté de l'égalité, on retrouve l'identité de convolution

$$\frac{1}{(\cosh\nu)^{1/2}} = \frac{1}{\pi}\int_0^\nu \frac{dx}{(\sinh x)^{1/2}\,(\sinh(\nu-x))^{1/2}\cosh(\nu-x)} \tag{8.89}$$

déjà discutée dans [116].

Remarque. Il n'est pas toujours évident, à partir d'une relation entre séries génératrices, d'en déduire une relation simple entre doubles transformées de Laplace de fonctionnelles quelconques. Ainsi, de la relation (8.7) liant chemins de Dyck et méandres, nous avons pu déduire la relation (8.41) entre doubles transformées de Laplace sur l'aire des excursions et méandres. Il semble néanmoins difficile d'en déduire une relation similaire pour des fonctionnelles plus générales.

8.5 Conclusion

L'utilisation des fonctions génératrices permet de retrouver plusieurs résultats sur l'aire du mouvement brownien, sans plus de prérequis que l'analyse de singularité, et donne une intuition sur les relations entre les différents types de distribution. Bien que le méthode des moments ne semble pas fournir de nouveaux résultats, elle peut être appliquée à des objets combinatoires qui ne convergent pas nécessairement vers un mouvement brownien connu.

Pour d'autres fonctionnelles du type x^2, il semble difficile de se débarrasser de variables catalytiques et d'obtenir la distribution limite.

Un travail futur possible comprend l'étude de la distribution jointe de l'aire et de l'altitude finales des méandres — comme effectuée dans [35] pour le mouvement brownien —, en utilisant des équations fonctionnelles du type [86]

$$F(z,u,v) = 1 + z\left(uv + \frac{1}{uv}\right)F(z,u,vu) - \frac{z}{uv}F(z,u,0), \tag{8.90}$$

où F est la fonction génératrice des méandres énumérés par taille (variable z), aire (variable u), et altitude finale (variable v). Pour ce type d'équation, la méthode des moments pourrait ne pas être suffisante.

Chapitre 9

Tries : occurrences de motifs et longueur de cheminement interne

Les tries sont des arbres permettant de ranger des mots ou clés engendrés de manière aléatoire. Pour une présentation de ces objets, nous renvoyons à [146] ou [36]. Contrairement aux arbres étudiés jusqu'ici dans cette thèse, on n'étudie habituellement des paramètres non pas en conditionnant par leur taille, mais par le nombre de mots qu'ils contiennent. En effet, même pour un nombre fini de mots, la taille du trie peut être arbitrairement grande, dans la mesure où des préfixes communs aux mots peuvent être arbitrairement longs. On étudie des paramètres limites sur des tries en faisant tendre vers l'infini le nombre de mots contenus dans le trie.

Des paramètres classiques [113, 114, 146, 169] étudiés sont la taille et la longueur de cheminement externe. La taille mesure l'espace mémoire occupé par le trie, tandis que la longueur de cheminement externe, divisée par le nombre de clés, compte le temps de recherche moyen d'une clé donnée.

Dans ce chapitre, nous considérons des tries formés sur des mots engendrés par une source de Bernoulli, c'est-à-dire que les lettres des mots sont des 0 ou des 1, et que chaque lettre est tirée indépendamment des autres suivant la distribution $\mathcal{B}(1,p)$.

Les tries font encore l'objet de recherches actuellement. On peut notamment considérer des sources de mots différentes, comme dans [38] ou des arbres proches des tries [37]. On notera également l'avènement des Patricia tries, qui sont des arbres obtenus à partir des tries, en compactant en un seul nœud les suites de nœuds internes à un seul fils, dont Jérémie Bourdon a étudié la taille et la longueur de cheminement [17, 18].

Notons enfin que les tries, avec des mots engendrés par une source de Bernoulli, sont étudiés actuellement par Michael Fuchs, Hsien-Kuei Hwang et Ralph Neininger, afin de classifier des distributions limites de paramètres définis récursivement sur des tries selon des fonctions de péage : un paramètre v défini sur un arbre t de sous-arbres t_1 et t_2 sera de la forme $v(t) = f(t) + v(t_1) + v(t_2)$, et selon ce que vaut f asymptotiquement, dans des échelles du type $O\big(n^\alpha \log \beta n\big)$, où n est le nombre de clés, on veut savoir quelle est la distribution limite de v. Cette volonté de classifier des distributions selon des fonctions de péage a déjà donné lieu à des études sur les arbres binaires de recherche [109] et sur les arbres de Catalan [65] ; dans les échelles asymptotiques du type $O\big(n^\alpha \log \beta n\big)$ concernant la fonction de péage, n correspond alors bien sûr à la taille n de l'arbre. Il est donc naturel de l'étendre aux tries.

Dans une mouvance similaire, ce chapitre cherche à généraliser aux tries des paramètres déjà étudiés et connus sur d'autres arbres. Nous chercherons, dans un premier temps, sur une suggestion de Julien Clément, à étudier le nombre d'occurrences d'un motif donné dans un trie.

Nous chercherons, dans un deuxième temps, à étudier la longueur de cheminement interne des tries.

Le nombre d'occurrences d'un motif donné dans un trie, même s'il n'a jamais été étudié auparavant, mène à des calculs pratiquement identiques à ceux qui ont déjà été faits pour la taille des tries, que ce soit par méthodes d'analyse complexe uniquement [113, 114, 146], ou en s'aidant également de méthodes de point fixe [169]. En outre, ce paramètre rentre dans les travaux en cours de Michael Fuchs, Hsien-Kuei Hwang et Ralph Neininger. Cette première partie constitue donc essentiellement une révision. On rappelle notamment comment faire usage de l'astuce consistant à poissoniser la source, rendant l'étude des moyenne et variance plus facile, par manipulations de séries génératrices et transformations de Mellin, puis à dépoissoniser la source, pour en déduire les résultats pour une source de Bernoulli. Au cours de cette révision, nous montrons le caractère gaussien de la distribution limite en utilisant seulement des méthodes d'analyse complexe, comme dans [113, 114, 146], puis des méthodes de point fixe, comme dans [169]. Nous couvrons le cas général $p \in [0, 1]$, où p est la probabilité d'avoir un 0. Notons que les méthodes de point fixe [169] ont seulement été présentées jusqu'à présent pour des sources de Bernoulli équilibrées, *i.e.* avec des 0 et 1 équiprobables, et que le cas général n'a pas encore donné lieu à publication. L'auteur de cette thèse n'a en outre pas accès aux travaux en cours de Michael Fuchs, Hsien-Kuei Hwang et Ralph Neininger.

La longueur de cheminement interne suit un phénomène différent, dans la mesure où il s'agit d'un paramètre récursif avec une fonction de péage égale à la taille du trie. En outre, la longueur de cheminement interne ne se déduit pas facilement de la longueur de cheminement externe comme dans le cas des arbres binaires [189, 188], car les tries sont des arbres unaires-binaires. Cette fonction de péage, même si elle est de moyenne linéaire, est arbitrairement grande, si bien que les travaux en cours de Michael Fuchs, Hsien-Kuei Hwang et Ralph Neininger ne couvrent pas l'étude de la longueur de cheminement interne. L'étude de la longueur de cheminement interne requiert en fait d'étudier la distribution jointe de la taille et de la longueur de cheminement interne. Il semble qu'on ne puisse plus se contenter d'analyse complexe pour déterminer la distribution limite, et cela implique notamment de développer des résultats intermédiaires pour montrer qu'on se trouve dans de bonnes conditions de dépoissonisation. Munis de la matrice de covariance de la taille et de la longueur de cheminement interne, nous montrons le caractère gaussien de la distribution limite jointe (taille, longueur de cheminement interne).

La longueur de cheminement interne possède un intérêt en elle-même dans la mesure où elles s'interprète physiquement comme la somme des longueurs des préfixes communs stricts des clés du trie. L'étude mathématique de ce trie procède aussi d'une volonté d'étudier des paramètres vus précédemment comme le moment d'inertie. On pourrait en utilisant la même méthode que dans ce chapitre étudier le moment d'inertie externe des tries et montrer qu'on a une limite gaussienne. Si on veut étudier des paramètres du type moment d'inertie interne ou somme des cubes des distances des nœuds à la racine, voire la somme des puissances de ces distances, il faudrait alors pouvoir gérer des matrices de taille au moins 3×3. On peut alors naturellement conjecturer des limites jointes gaussiennes entre tous ces paramètres, mais il n'est pas aisé d'étendre le calcul car ici on utilise une astuce de calcul pour calculer la racine carrée nos matrices de taille 2×2, due à Bernard W. Levinger [138].

9.1 Rappels

Nous effectuons des rappels sur les zéros de $1 - p^{-s} - q^{-s}$, sur la transformation de Mellin, et sur la méthode de poissonisation. Nous établissons quelques lemmes utiles permettant d'établir

des conditions de dépoissonisation.

9.1.1 Zéros de $1 - p^{-s} - q^{-s}$

Lors de l'inversion de la transformation de Mellin, on est amené à effectuer un calcul de résidus. Or, ces résidus sont souvent issus de pôles de fractions du type

$$\frac{1}{1 - p^{-z} - q^{-z}},$$

d'où l'utilité du lemme ci-dessous.

Lemme 9.1 (Jacquet 1989, Schachinger 1993)
(o) *Les solutions de*

$$p^{-z} + q^{-z} = 1,$$

pour $p + q = 1$, p, q réels et z complexe, sont en nombre dénombrable. Soient z_k, $k \in \mathbb{Z}$, ces solutions.

(i) *Pour tout $k \in \mathbb{Z}$,*

$$-1 \leqslant \mathrm{Re}(z_k) \leqslant \sigma_0,$$

où σ_0 est la solution positive de $1 + q^{-s} = p^{-s}$. En outre,

$$\frac{(2k-1)\pi}{\log p} \leqslant \mathrm{Im}(z_k) \leqslant \frac{(2k+1)\pi}{\log p}.$$

(ii) *Si $\mathrm{Re}(z_k) = -1$ et $\mathrm{Im}(z_k) \neq 0$, alors $\log p / \log q$ est rationnel. Dans ce cas, plus précisément, si $\dfrac{\log p}{\log q} = \dfrac{s_1}{s_2}$, où $\mathrm{pgcd}(s_1, s_2) = 1$ pour $s_1, s_2 \in \mathbb{Z}$, alors*

$$z_k = -1 + \frac{2ks_1\pi i}{\log p}$$

pour tout $k \in \mathbb{Z}$.

En particulier, on a la conséquence suivante :

Corollaire 9.1 *Pour tout k,*

$$|z_k|^2 \leqslant \max(|-1|, \sigma_0)^2 + \left(\frac{(2k+1)\pi}{\log p}\right)^2.$$

Le lemme précédent dit donc que, dans le cas $\dfrac{\log p}{\log q}$ irrationnel, les racines de $1 - p^{-s} - q^{-s}$ différentes de -1 possèdent une partie réelle strictement plus grande que -1.

On peut décrire les racines de $1 - p^{-s} - q^{-s}$ dans le cas $\dfrac{\log p}{\log q}$ rationnel de manière plus fine :

Lemme 9.2 *Supposons $\dfrac{\log p}{\log q} \in \mathbb{Q}$, de la forme $\dfrac{\log p}{\log q} = \dfrac{s_1}{s_2}$. L'ensemble des parties réelles des racines de $1 - p^{-s} - q^{-s}$ est fini. Soient r_1, r_2, ..., r_M ces réels. Soit $\mathcal{S}$ l'ensemble des racines de $1 - p^{-s} - q^{-s}$. Alors il existe z_1, z_2, ..., z_M dans $\mathcal{S}$, de parties réelles r_1, r_2, ..., r_M, tels que*

$$\mathcal{S} = \bigcup_{j=1}^{m} \left\{ z_j \cdot \exp\left(2\pi i \frac{ks_1}{\log p}\right) \;\middle|\; k \in \mathbb{Z} \right\}.$$

Démonstration. Posons

$$\alpha = \frac{\log p}{\log q} = \frac{s_1}{s_2}.$$

Il vient

$$\begin{aligned} p^{-s} + q^{-s} &= \exp(-s\log p) + \exp(-s\log q) \\ &= (e^{-\alpha s})^{s_1} + (e^{-\alpha s})^{s_2}. \end{aligned}$$

Donc s est solution de $p^{-s} + q^{-s} = 1$ si et seulement si $e^{-\alpha s}$ est solution de $X^{s_1} + X^{s_2} = 1$.

Maintenant, à chaque racine x du polynôme $X^{s_1} + X^{s_2} - 1$ correspond une infinité de solutions s de $p^{-s} + q^{-s} = 1$ de la forme

$$s = -\frac{\log x}{\alpha} + 2\pi i \frac{k}{\alpha} = -\frac{\log x}{\alpha} + 2\pi i \frac{k s_1}{\log p}. \qquad (k \in \mathbb{Z})$$

□

Corollaire 9.2 *Soit $S(x)$ la somme des résidus de $\frac{a(s)x^{-s}}{1-p^{-s}-q^{-s}}$, avec $|a(s+ik)| \underset{k\to\infty}{\longrightarrow} 0$ exponentiellement en k — typiquement une fonction Γ — et $a(z_k) \neq 0$, où les z_k sont les zéros de $1 - p^{-s} - q^{-s}$. Alors si $\frac{\log p}{\log q} \in \mathbb{Q}$, $\log_\gamma x \mapsto S(x)$, avec γ tel que*

$$\log_\gamma p = \frac{\log p}{-\frac{\log p}{s_1}} = -s_1, \ \log_\gamma q = -s_2,$$

est une fonction 1-périodique de $\log_\gamma x$. Si $\frac{\log p}{\log q} \notin \mathbb{Q}$, alors S est de la forme

$$S(x) = \frac{a(-1)}{p\log p + q\log q} x^{-1}(1 + o(x))$$

pour x tendant vers $+\infty$.

Démonstration. S est une somme infinie de $\lambda_k e^{-\frac{2ks_1\pi i}{\log p}}$. Soit la fonction

$$h(t) = p^{-t}\log p + q^{-t}\log q,$$

et par abus on notera $h = h(-1)$. On a

$$S(x) = \sum_{k\in\mathbb{Z}} \frac{a(s_k)}{h(z_k)} x^{-z_k}.$$

Si $\frac{\log p}{\log q} \in \mathbb{Q}$, alors pour tout k, d'après la forme générale des z_k, on voit facilement que $\log_\gamma x \mapsto x^{-z_k}$ est une fonction 1-périodique de $\log_\gamma x$. Il en est de même de $\log_\gamma x \mapsto S(x)$.

Si $\frac{\log p}{\log q} \notin \mathbb{Q}$, alors

$$S(x) = \frac{a(-1)}{h(-1)} x^{-1} + \sum_{k\in\mathbb{Z}^*} \frac{a(z_k)}{h(z_k)} x^{-z_k}.$$

Montrons que $\sum_{k\in\mathbb{Z}^*} \frac{a(z_k)}{h(z_k)} x^{-z_k}$ tend vers zéro quand x tend vers $+\infty$, avec $\arg(x) < \theta$. Soit $\varepsilon > 0$. Soit $K \in \mathbb{N}^*$ tel que

$$\sum_{|k|>K} \left|\frac{a(z_k)}{h(z_k)}\right| < \varepsilon.$$

Soit alors $A > 1$ tel que pour $|x| > A$,

$$\sum_{|k|\leqslant K} \left|\frac{a(z_k)}{h(z_k)}\right| \left|x^{-z_k}\right| < \varepsilon.$$

Alors pour $|x| > A$,

$$\sum_{|k|\in\mathbb{Z}^*} \left|\frac{a(z_k)}{h(z_k)}\right| \left|x^{-z_k}\right| < 2\varepsilon.$$

□

Si on s'intéresse plutôt à la somme S des résidus de $\frac{a(s)x^{-s}}{(1-p^{-s}-q^{-s})^m}$, alors on a aussi une somme de terme de la forme $x^{-z_k}\log^d x$, et on peut faire une discussion similaire.

9.1.2 Transformées de Mellin

Définition 9.1 *Soit f une fonction définie sur $\mathbb{R}_+$. Sa transformée de Mellin est donnée par*

$$\mathcal{M}\{f(x);s\} := f^*(s) = \int_0^\infty f(x)x^{s-1}\,dx.$$

On trouve dans [195] les propositions suivantes :

Proposition 9.1 *Soit f une fonction deux fois différentiable. Si sa transformée de Mellin $f^*(s)$ est de la forme*

$$f^*(s) = \sum_{k=0}^{K} \frac{d_k}{(s-b)^{k+1}},$$

et si $f^*(s)$ peut être analytiquement prolongée à $\beta \leqslant \mathrm{Re}(s) < M$, alors

$$f(x) \underset{x\to\infty}{=} -\sum_{k=0}^{K}(-1)^k \frac{d_k}{k!} x^{-b}\log^k x + O\left(x^{-M}\right).$$

Proposition 9.2 *Soit f une fonction continue de transformée de Mellin $f^*(s)$ définie sur une bande fondamentale non vide $\langle\alpha,\beta\rangle$. Si f admet le développement asymptotique*

$$f(x) \underset{x\to\infty}{=} -\sum_{k=0}^{K}(-1)^k \frac{d_k}{k!} x^{-b}\log^k x + O\left(x^{-M}\right),$$

alors

$$f^*(s) \asymp -\sum_{k,b} c_{k,b}\frac{(-1)^k k!}{(s-b)^{k+1}}$$

pour $\alpha < \mathrm{Re}(s) < M$.

9.1.3 Dépoissonisation

Définition 9.2 *Soit* $\theta \in \left]0, \frac{\pi}{2}\right[$. *Alors* $\mathcal{S}_\theta = \{z \mid |\arg(z)| \leqslant \theta\}$.

Théorème 9.1 *Supposons les deux conditions suivantes vérifiées pour des nombres* A, B, $R > 0$ *et* $\alpha > 0$, β *et* γ :
(I) Pour $z \in \mathcal{S}_\theta$,

$$|z| > R \implies \left|\widetilde{G}(z)\right| \leqslant B|z|^\beta \Psi(|z|),$$

où Ψ *est une fonction à faible variation,* i.e. *telle que pour* t *fixé,* $\lim_{x\to\infty} \frac{\Psi(tx)}{\Psi(x)} = 1$ (e.g. $\Psi(x) = \log^d x$ *pour un* $d > 0$) ;
(O) Pour tout $z = \rho e^{i\vartheta}$ *avec* $\vartheta \leqslant \pi$ *tel que* $z \notin \mathcal{S}_\theta$ *et*

$$\rho = |z| > R \;\to\; \left|\widetilde{G}(z)e^z\right| \leqslant A\rho^\gamma \exp\left[(1 - \alpha\vartheta^2)\rho\right].$$

Alors pour tout entier m *positif*

$$\begin{aligned} g_n &= \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij} n^i \widetilde{G}^{(j)}(n) + O\left(n^{\beta-m-1}\Psi(n)\right) \\ &= \widetilde{G}(n) + \sum_{k=1}^{m}\sum_{i=1}^{k} b_{i,k+i} n^i \widetilde{G}^{(k+i)}(n) + O\left(n^{\beta-m-1}\Psi(n)\right), \end{aligned}$$

où les b_{ij} *sont les coefficients de* $\exp(x\log(1+y) - xy)$ *en* $x^i y^j$, *soit*

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} b_{ij} x^i y^j = \exp(x\log(1+y) - xy)$$

avec $b_{ij} = 0$ *pour* $j < 2i$.

Les premiers termes du développement ci-dessus s'écrivent

$$g_n = \widetilde{G}(n) - \frac{1}{2}n\widetilde{G}^{(2)}(n) + \frac{1}{3}n\widetilde{G}^{(3)}(n) + \frac{1}{3}n\widetilde{G}^{(3)}(n).$$

Théorème 9.2 *Soient* $\widetilde{X}(z)$ *et* $\widetilde{V}(z) + \widetilde{X}^2(z)$ *satisfaisant la condition (O) et* $\widetilde{X}(z)$ *et* $\widetilde{V}(z)$ *satisfaisant la condition (I) du Théorème 9.1 avec* $\beta \leqslant 1$, e.g. $\widetilde{X}(z) = O\left(z^\beta\Psi(z)\right)$, *et* $\widetilde{V}(z) = O\left(z^\beta\Psi(z)\right)$ *dans un cône linéaire* $\mathcal{S}_\theta$ *et les conditions appropriées (O) en dehors du cône, où* $\Psi(z)$ *est une fonction à faible variation. Alors les égalités*

$$\begin{aligned} \mathbf{E}[X_n] &= \widetilde{X}(n) - \frac{1}{2}n\widetilde{X}^{(2)}(n) + O\big(n^{\beta-2}\Psi(n)\big), \\ \mathbf{Var}[X_n] &= \widetilde{V}(n) - n\big[\widetilde{X}'_n\big]^2 + O\big(n^{\beta-1}\Psi(n)\big), \end{aligned}$$

sont vérifiées pour n *grand.*

Théorème 9.3 *Si* $\widetilde{X}(z) = O\left(z^\beta\Psi(z)\right)$, $\widetilde{Y}(z) = O\left(z^\beta\Psi(z)\right)$, $\widetilde{C}(z) = O\left(z^\beta\Psi(z)\right)$, *dans un cône linéaire* $\mathcal{S}_\theta$ *et si* $\widetilde{X}(z)$, $\widetilde{Y}(z)$ *et* $\widetilde{C}(z) + \widetilde{X}(z)\widetilde{Y}(z)$ *satisfont la condition (O) en dehors du cône, alors la covariance dépoissonisée s'écrit :*

$$E[X_nY_n] - E[X_n]E[Y_n] = \widetilde{C}(n) - n\widetilde{X}'(n)\widetilde{Y}'(n) + O\left(n^{\beta-1}\Psi(n)\right).$$

Démonstration. On a

$$\begin{aligned}
\widetilde{G}(z) &= \widetilde{C}(z) + \widetilde{X}(z)\widetilde{Y}(z), \\
\widetilde{G}'(z) &= \widetilde{C}'(z) + \widetilde{X}'(z)\widetilde{Y}(z), + \widetilde{X}(z)\widetilde{Y}'(z), \\
\widetilde{G}^{(2)}(z) &= \widetilde{C}(z) + \widetilde{X}^{(2)}(z)\widetilde{Y}(z) + \widetilde{X}'(z)\widetilde{Y}'(z) + \widetilde{X}(z)\widetilde{Y}^{(2)}(z), \\
E[X_n Y_n] &= \widetilde{C}(n) + \widetilde{X}(n)\widetilde{Y}(n) - \frac{1}{2}n\Big(\widetilde{C}^{(2)}(n) + O\left(n^{\beta-2}\Psi(n)\right), \\
&\quad + \widetilde{X}^{(2)}(n)\widetilde{Y}(n) + \widetilde{X}(n)\widetilde{Y}^{(2)}(n) + 2\widetilde{X}'(n)\widetilde{Y}'(n)\Big), \\
&\quad + O\left(n^{\beta-1}\Psi(n)\right), \\
E[X_n Y_n] - E[X_n]E[Y_n] &= \widetilde{C}(n) - n\widetilde{X}'(n)\widetilde{Y}'(n) + O\left(n^{\beta-1}\Psi(n)\right),
\end{aligned}$$

car $n\widetilde{C}^{(2)}(n) = O\left(n^{\beta-1}\Psi(n)\right)$. □

Dans [195], la preuve du théorème 10.6 permet d'affirmer

Théorème 9.4 *Considérons l'équation fonctionnelle suivante, avec $p, q > 0$ de somme 1,*

$$\widetilde{G}(z) = \gamma_1(z)\widetilde{G}(zp) + \gamma_2(z)\widetilde{G}(zq) + t(z),$$

et supposons qu'elle ait une solution $\widetilde{G}$ entière. Supposons en outre qu'il existe $\beta > 0$, $0 < \theta < \pi/2$, et $0 < \mu < 1$ tels que pour $|z| > \xi$:

(I) Pour $z \in \mathcal{S}_\theta$

$$\begin{aligned}
|\gamma_1(z)|\, p^\beta + |\gamma_2(z)|\, q^\beta &\leqslant 1 - \mu, \\
|t(z)| &\leqslant B\mu|z|^\beta;
\end{aligned}$$

(O) Pour $z \notin \mathcal{S}_\theta$ et pour $\alpha < 1$,

$$\begin{aligned}
|\gamma_1(z)|\, e^{q\mathrm{Re}(z)} &\leqslant \frac{1}{3}e^{\alpha|z|q}, \\
|\gamma_1(z)|\, e^{p\mathrm{Re}(z)} &\leqslant \frac{1}{3}e^{\alpha|z|p}, \\
|t(z)|\, e^{\mathrm{Re}(z)} &\leqslant \frac{1}{3}e^{\alpha|z|}.
\end{aligned}$$

Alors il existe $B, B' > 0$, et $\alpha < 1$ tels que

$$\forall z \in \mathcal{S}_\theta,\ \left|\widetilde{G}(z)\right| \leqslant B|z|^\beta,$$

et

$$\forall z \notin \mathcal{S}_\theta,\ \left|\widetilde{G}(z)e^z\right| \leqslant B'|z|^{\alpha|z|}.$$

Corollaire 9.3 *Sous les mêmes hypothèses, on a*

$$g_n = \widetilde{G}(n) + O\left(n^{\beta-1}\right),$$

avec $g_n = [z^n]\left(n!e^z\widetilde{G}(z)\right)$.

Remarque 9.1 *Quitte à augmenter α et ξ, on peut choisir B' arbitrairement petit.*

Lemme 9.3 *Soient G et t deux séries entières de rayon de convergence infini telles que :*

$$\widetilde{G}(z) = \widetilde{G}(pz) + \widetilde{G}(qz) + t(z),$$

avec $p+q=1$ et $p>0$, $q>0$. Supposons qu'il existe des nombres A, B, $R>0$ et $\alpha>0$, et γ :
(O) Pour tout $z=\rho e^{i\vartheta}$ avec $\theta \leqslant \pi$ tel que $z \notin \mathcal{S}_\theta$ et

$$\rho = |z| > R \implies |t(z)e^z| \leqslant A\rho^\gamma \exp\left[(1-\alpha\vartheta^2)\rho\right].$$

Alors il existe des nombres A', B', $R'>0$ et $\alpha'>0$, et γ' :
(O') Pour tout $z=\rho e^{i\vartheta}$ avec $\theta \leqslant \pi$ tel que $z \notin \mathcal{S}_\theta$ et

$$\rho = |z| > R' \implies \left|\widetilde{G}(z)e^z\right| \leqslant A'\rho^{\gamma'} \exp\left[(1-\alpha'\vartheta^2)\rho\right].$$

Démonstration. Soit $G(z) = \widetilde{G}(z)e^z$. Alors G vérifie

$$G(z) = G(pz)e^{qz} + G(qz)e^{pz} + t(z)e^z.$$

Soient $\gamma' > \max(\gamma, 1)$ et R' tel que pour tout $\vartheta \in [\theta, \pi]$

$$\rho > R' \implies A\rho^\gamma \exp\left[(1-\alpha\vartheta^2)\rho\right] \leqslant (1-p^{\gamma'}-q^{\gamma'}) \exp\left[(1-\alpha\vartheta^2)\rho\right].$$

Supposons $p \leqslant q$. Soient $A' > \max(A,1)$ et $\alpha' < \min(\alpha, \frac{1}{2})$ tel que

$$\left|\widetilde{G}(z)e^z\right| \leqslant A'\rho^{\gamma'} \exp\left[(1-\alpha'\vartheta^2)\rho\right]$$

pour tout $z=\rho e^{i\vartheta}$ tel que $z \notin \mathcal{S}_\theta$ et $R' < \rho \leqslant R'/p$. Posons

$$\mathcal{R}_m = \left\{ z \notin \mathcal{S}_\theta,\ R' < |z| \leqslant \frac{R'}{pq^{m+1}} \right\}.$$

Il est clair que (O') est vérifié pour $z \in \mathcal{R}_0$. Montrons par récurrence sur m que pour tout m, si $z \in \mathcal{R}_m$, alors (O') est vérifié. On remarque d'abord la propriété

$$z \in \mathcal{R}_{m+1} \setminus \mathcal{R}_0 \implies pz \in \mathcal{R}_m \text{ et } qz \in \mathcal{R}_m.$$

Il s'ensuit donc, pour $z=\rho e^{i\vartheta} \in \mathcal{R}_{m+1} \setminus \mathcal{R}_0$,

$$\begin{aligned}
|G(z)| &\leqslant |G(pz)| \cdot |e^{qz}| + |G(qz)| \cdot |e^{pz}| + |t(z)e^z| \\
&\leqslant A'(p\rho)^{\gamma'} \exp\left[(1-\alpha'\vartheta^2)p\rho\right] \exp\left[(1-\alpha'\vartheta^2)q\rho\right] \\
&\quad + A'(q\rho)^{\gamma'} \exp\left[(1-\alpha'\vartheta^2)q\rho\right] \exp\left[(1-\alpha'\vartheta^2)p\rho\right] \\
&\quad + (1-p^{\gamma'}-q^{\gamma'}) \exp\left[(1-\alpha'\vartheta^2)\rho\right] \\
&\leqslant A'\rho^{\gamma'} \exp\left[(1-\alpha'\vartheta^2)\rho\right].
\end{aligned}$$

□

Lemme 9.4 *Soient $p<q$ deux réels strictement positifs de somme 1. Soient $a(z)$ et $r(z)$ deux séries entières analytiques dans tout $\mathcal{C}$ telles que*

$$a(z) = a(pz) + a(qz) + r(z),$$

$$a(0) = 0,$$

et

$$r(z) = O(|z|^{\beta})$$

pour $|z|$ *tendant vers* ∞, z *restant dans un cône* $\mathcal{S}_\theta$ *avec* $\beta \geqslant 1$. *Alors pour tout* $\varepsilon > 0$,

$$a(z) = O(|z|^{\beta+\varepsilon}).$$

Démonstration. Soient $\varepsilon > 0$, et $b(z) = \frac{a(z)}{z^{\beta+\varepsilon}}$. Posons $\beta' = \beta + \varepsilon$. On a

$$b(z) = p^{\beta'} b(pz) + q^{\beta'} b(z) + \frac{r(z)}{z^{\beta'}}.$$

Or

$$r(z) = O(|z|^{\beta}),$$

donc

$$\frac{r(z)}{z^{\beta'}} = O(|z|^{\varepsilon}).$$

Soient $A > 0$ et M tels que $M > \sup_{\mathrm{Re}(z)\leqslant A, z\in\mathcal{S}_\theta} |b(z)|$, et que pour $\mathrm{Re}(z) > A$ ($z \in \mathcal{S}_\theta$), on ait

$$\left|\frac{r(z)}{z^{\beta'}}\right| < (1 - p^{\beta'} - q^{\beta'})M.$$

Soient $\mathcal{C}_m$ les portions de cônes linéaires définis par

$$\mathcal{C}_m = \left\{ z \in \mathcal{S}_\theta,\ \mathrm{Re}(z) \leqslant \frac{A}{q^m} \right\}.$$

Montrons par récurrence sur m que $\sup_{z\in\mathcal{C}_m} |b(z)|$ est majoré par M. C'est vrai au rang $m = 0$. Supposons la propriété vérifiée au rang m. Soit $z \in \mathcal{C}_{m+1}$. On a $pz,\ qz \in \mathcal{C}_m$, donc

$$|b(z)| \leqslant p^{\beta'} |b(pz)| + q^{\beta'} |b(z)| + \left|\frac{r(z)}{z^{\beta'}}\right| \leqslant p^{\beta'} M + q^{\beta'} M + \left(1 - p^{\beta'} - q^{\beta'}\right) M \leqslant M.$$

On en conclut

$$\forall z \in \mathcal{S}_\theta,\ |a(z)| \leqslant M|z|^{\beta+\varepsilon}.$$

□

9.2 Occurrences de motifs

Cette partie est inspirée de [88, 73], sur une suggestion de Julien Clément.

On cherche le nombre d'occurrences d'un trie $\mathfrak{m}$ fixé à m clés dans les tries, indépendamment des étiquettes des tries. On tire les 0 et les 1 indépendamment avec probabilités respectives p et $q = 1 - p$ (loi de Bernoulli $\mathcal{B}(1, p)$). En particulier, les tailles $I_1^{(n)}$ et $I_2^{(n)}$ des sous-arbres gauche et droit d'un trie à n clés suivent respectivement les lois $\mathcal{B}(n, p)$ et $\mathcal{B}(n, q)$.

Mathématiquement parlant, cette section n'est donnée qu'à titre pédagogique, notamment par méthode de point fixe, puisque l'étude est la même que pour la taille. En outre, cette étude rentre dans le cadre des travaux — encore en cours de rédaction — de Michael Fuchs, Hsien-Kuei Hwang et Ralph Neininger, partiellement présentés au *Ninth Seminar on Analysis of Algorithms*,

qui incluront aussi le cas $p \neq q$, avec une preuve sans doute similaire à celle qui est présentée ici. L'article [169] n'étudie la taille avec des méthodes de point fixe que dans le cas $p = q = 1/2$.

Néanmoins, il est clair qu'on peut se cantonner aux cas $m \geqslant 2$: pour $m = 1$, $\mathfrak{m}$ serait un nœud externe, cas trivial, et pour $m = 0$, $\mathfrak{m}$ serait un nœud interne, ce qui revient à l'étude déjà faite de la taille d'un trie.

Théorème 9.5 *Soit $\mathfrak{m}$ un arbre fixé à m nœuds externes, avec $m \geqslant 2$. Soit $\omega[t]$ la variable aléatoire comptant le nombre d'occurrences de $\mathfrak{m}$ dans t. Soit*

$$T(z,u) = \sum_t \sum_{k\geqslant 0} P[\omega_n = k] p_t z^{|t|} u^k$$

la série génératrice des tries comptés par nombre de nœuds externes selon z dans le modèle de Poisson et par nombre d'occurrences de $\mathfrak{m}$ dans les tries selon u. Alors

$$T(z,u) = T(pz,u)\,T(qz,u) + p_{\mathfrak{m}} \frac{z^m}{m!}(u-1)e^{-z}. \tag{9.1}$$

En outre, T est analytique pour tous z et u.

Démonstration. Soit $[\![t = \mathfrak{m}]\!]$ la variable aléatoire indicatrice de l'événement $t = \mathfrak{m}$. Alors

$$\omega[t] = \omega[t_0] + \omega[t_1] + [\![t = \mathfrak{m}]\!],\ \omega[\emptyset] = 0.$$

Soient t_0 et t_1 les sous-arbres gauche et droit de t. Il vient alors

$$u^{\omega[t]} = u^{[\![t=\mathfrak{m}]\!]} u^{\omega[t_0]} u^{\omega[t_1]} = u^{\omega[t_0]} u^{\omega[t_1]} + [\![t = \mathfrak{m}]\!](u-1).$$

En effet, $[\![t = \mathfrak{m}]\!] = 1$ entraîne $\omega[t_0] = \omega[t_1] = 0$.

Soient $\mathfrak{m}$ un arbre fixé à $m \geq 2$ clés, et ω_n le nombre d'occurrences de $\mathfrak{m}$ dans les arbres à n clés. On notera p_t la probabilité de l'arbre t, définie récursivement par

$$p_t = p^{|t_1|} q^{|t_2|} \binom{|t|}{|t_1|} p_{t_1} p_{t_2}$$

avec $|t|$ égal au nombre de clés de t, et $t = \circ(t_1, t_2)$. Alors

$$\begin{aligned}
P[\omega_n = k | N_1 = n_1, N_2 = n_2] &= \sum_{k_1+k_2=k} P[\omega_{n1} = k_1] P[\omega_{n2} = k_2] + p_{\mathfrak{m}}, \\
\sum_{\substack{n\geqslant 2\\ k\geqslant 0}} P[\omega_n = k] \frac{z^n}{n!} u^k e^{-z} &= \sum_{\substack{n\geqslant 2\\ k\geqslant 0}} \sum_{\substack{k_1+k_2=k\\ n_1+n_2=n}} p^{n_1} \frac{P[\omega_{n1} = k_1]}{n_1!} q^{n_2} \frac{P[\omega_{n2} = k_2]}{n_2!} z^n u^k e^{-z} \\
&\quad + p_{\mathfrak{m}} \frac{z^m}{m!}(u-1)e^{-z}, \\
T(z,u) - e^{-z} - ze^{-z} &= T(pz,u)\,T(qz,u) - e^{-z} - ze^{-z} + p_{\mathfrak{m}} \frac{z^m}{m!}(u-1)e^{-z}, \\
T(z,u) &= T(pz,u)\,T(qz,u) + p_{\mathfrak{m}} \frac{z^m}{m!}(u-1)e^{-z},
\end{aligned}$$

car

$$\begin{aligned}
&\sum_{\substack{n<2\\ k\geqslant 0}} \sum_{\substack{k_1+k_2=k\\ n_1+n_2=n}} p^{n_1} \frac{P[\omega_{n1} = k_1]}{n_1!} q^{n_2} \frac{P[\omega_{n2} = k_2]}{n_2!} z^n u^k e^{-z} \\
&= \sum_{n<2} \sum_{n_1+n_2=n} p^{n_1} \frac{P[\omega_{n1} = 0]}{n_1!} q^{n_2} \frac{P[\omega_{n2} = 0]}{n_2!} z^n u^0 e^{-z} \\
&= z^0 e^{-z} + (p \cdot 1 + 1 \cdot q)\, z e^{-z} \\
&= e^{-z} + z e^{-z}.
\end{aligned}$$

Le nombre d'occurrences de $\mathfrak{m}$ dans un arbre t étant inférieur au nombre de clés, $P(z,u)$ est clairement analytique pour tous z et u. □

9.2.1 Moyenne et variance

Nous évaluons l'asymptotique de la moyenne $X_B(n)$ et la variance $V_B(n)$ du nombre d'occurrences d'un motif dans le modèle de Bernoulli. Nous devons au passage évaluer l'asymptotique des grandeurs similaires $X(z)$ et $V(z)$, puis dépoissoniser $X(z)$ et $V(z)+X^2(z)$.

Moyenne

Dans cette section, nous montrons que la moyenne du nombre d'occurrences d'un motif est linéaire en le nombre de clés, à une fonction périodique près.

Théorème 9.6 *Soit ω_n la variable aléatoire comptant le nombre d'occurrences de l'arbre $\mathfrak{m}$ dans un trie à n nœuds externes tiré au hasard. Alors la moyenne de ω_n est asymptotiquement linéaire, à une fonction périodique près de petite amplitude si $\log p/\log q \in \mathbb{Q}$. Plus précisément,*

$$\mathbf{E}[\omega_n] = \frac{p_{\mathfrak{m}}}{m(m-1)\left(\frac{1}{p}\log p + \frac{1}{q}\log q\right)} nG_1(n) + O(1),$$

où G_1 est une fonction de propriétés différentes suivant la rationalité de $\log p/\log q$.

Si $\log p/\log q \notin \mathbb{Q}$, alors G_1 est de la forme $1+o(n)$.

Si $\log p/\log q = s/t \in \mathbb{Q}$, on peut écrire G_1 comme combinaison linéaire de fonctions de la forme $n^{1-a_k}\varpi_k(\log_\gamma n)$, avec γ vérifiant

$$\log_\gamma p = \frac{\log p}{-\frac{\log p}{s}_1} = -s_1,\ \log_\gamma q = -s_2,$$

les a_k étant les différentes parties réelles possibles des racines de $1-p^{-s}-q^{-s}$, et les fonctions ϖ_k étant périodiques de période 1, de moyenne 1, et d'amplitude négligeable par rapport à 1.

Démonstration. En dérivant (9.1) par rapport à u et en faisant $u=1$ (rappelons que $T(z,1)=1$) :

$$\begin{aligned} T_u(z,1) &= T_u(pz,1) + T_u(qz,1) + p_{\mathfrak{m}}\frac{z^m}{m!}e^{-z} \\ X(z) &= X(pz) + X(qz) + p_{\mathfrak{m}}\frac{z^m}{m!}e^{-z} \end{aligned}$$

Donc la transformée de Mellin $X^*(s) = \mathcal{M}(X;s)$ vérifie :

$$\begin{aligned} X^*(s) &= \left(p^{-s}+q^{-s}\right)X^*(s) + p_{\mathfrak{m}}\mathcal{M}\left(\frac{z^m}{m!}e^{-z};s\right) \\ &= \left(p^{-s}+q^{-s}\right)X^*(s) + p_{\mathfrak{m}}\int_0^\infty \frac{x^m}{m!}e^{-x}x^{s-1}\,dx \\ &= \left(p^{-s}+q^{-s}\right)X^*(s) + \frac{p_{\mathfrak{m}}}{m!}\Gamma(m+s) \\ &= \frac{p_{\mathfrak{m}}\Gamma(m+s)}{m!\left(1-p^{-s}-q^{-s}\right)}. \end{aligned}$$

Comme $m \geqslant 2$, on a $\omega_n = 0$ pour $n \leqslant 1$, donc $X(x) \underset{x\to 0}{=} O\left(x^2\right)$. En outre, on a

$$X(x) \underset{x\to\infty}{=} O\left(x^{1+\varepsilon}\right)$$

pour tout $\varepsilon > 0$, d'après le lemme 9.4, si bien que X^* admet $\langle -2, -1\rangle$ comme bande fondamentale, à l'intérieur de laquelle seuls les résidus z_k sont à prendre en compte. Donc, avec $-2 < c < 1$ et $d > 0$:

$$\begin{aligned} X(x) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{p_{\mathrm{m}}\Gamma(s+m)}{m!\,(1-p^{-s}-q^{-s})} x^{-s}\, ds \\ &= O(x^{-d}) - \sum_{k\in\mathbb{Z}} \operatorname*{Res}_{s=z_k} \frac{p_{\mathrm{m}}\Gamma(s+m)}{m!\,(1-p^{-s}-q^{-s})} x^{-s} \\ &= O(x^{-d}) + \sum_{k\in\mathbb{Z}} \frac{p_{\mathrm{m}}\Gamma(z_k+m)}{m! h(z_k)} x^{-z_k}. \end{aligned}$$

On obtient

$$X(x) = \frac{p_{\mathrm{m}}}{m(m-1)\left(\frac{1}{p}\log p + \frac{1}{q}\log q\right)} x G_1(x) + O(x^{-d}), \tag{9.2}$$

où G_1 vérifie les propriétés énoncées suivant la rationalité de $\log p/\log q$: dans le cas irrationnel, on utlise le corollaire 9.2 ; dans le cas irrationnel, on utlise le le lemme 9.2 et on regroupe les racines qui sont conjuguées.

On trouve après dépoissonisation que $\mathbf{E}[\omega_n]$ est bien de la forme indiquée.

□

On peut obtenir une expression plus explicite dans le cas $p = q = 1/2$. On connaît en effet tous les résidus :

- $z_k = -1 + 2\pi i k/\log 2$, $\mathbb{Z}$, de $2^{s+1} - 1$.
- $s = -m - j$, $j \in \mathbb{N}^*$, de $\Gamma(m+s)$.

On a alors

$$\begin{aligned} X(x) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{p_{\mathrm{m}}\Gamma(s+m)}{m!\,(1-2^{s+1})} x^{-s}\, ds \\ &= O(x^{-d}) - \sum_{k\in\mathbb{Z}} \operatorname*{Res}_{s=z_k} \frac{p_{\mathrm{m}}\Gamma(s+m)}{m!\,(1-2^{s+1})} x^{-s} \\ &= O(x^{-d}) + \sum_{k\in\mathbb{Z}} \frac{p_{\mathrm{m}}\Gamma(z_k+m)}{m!\log 2} x^{-z_k} \\ &= O(x^{-d}) + \frac{p_{\mathrm{m}}\Gamma(m-1)}{m!\log 2} x + \sum_{k\in\mathbb{Z}^*} \frac{p_{\mathrm{m}}\Gamma(m-1+\frac{2\pi i k}{\log 2})}{m!\log 2} x^{1-\frac{2\pi i k}{\log 2}} \\ X(x) &= O(x^{-d}) + \left(\frac{p_{\mathrm{m}}}{m(m-1)\log 2} + \sum_{k\in\mathbb{Z}^*} \frac{p_{\mathrm{m}}\Gamma(m-1+\frac{2\pi i k}{\log 2})}{m!\log 2} e^{-\frac{2\pi i k}{\log 2}\log x}\right) x \\ X(x) &= O(x^{-d}) + \left(\frac{p_{\mathrm{m}}}{m(m-1)\log 2} + \sum_{k\in\mathbb{Z}^*} \frac{p_{\mathrm{m}}\Gamma(m-1+\frac{2\pi i k}{\log 2})}{m!\log 2} e^{-2\pi i k \log_2 x}\right) x \\ X(x) &= O(x^{-d}) + \frac{p_{\mathrm{m}}}{m(m-1)\log 2} \varpi_1(\log_2 x)\, x \end{aligned}$$

avec $\varpi_1(\log_2 x) = 1+\sum_{k\in\mathbb{Z}^*} \frac{\Gamma(m-1+\frac{2\pi ik}{\log 2})}{(m-2)!} e^{-2\pi ik\log_2 x}$. En dépoissonisant, on trouve que la moyenne $X(n)$ de ω_n vérifie asymptotiquement

$$X(n) = \frac{p_{\mathtt{m}}}{m(m-1)\log 2} n\varpi_1(\log_2 n) + O(1).$$

Variance

Dans cette section, nous montrons que la variance du nombre d'occurrences d'un motif est linéaire en le nombre de clés, à une fonction périodique près.

Théorème 9.7 *Soit ω_n la variable aléatoire comptant le nombre d'occurrences de l'arbre $\mathtt{m}$ dans un trie à n nœuds externes tiré au hasard. Alors la variance de ω_n est asymptotiquement linéaire, à une fonction périodique près de petite amplitude si $\log p/\log q \in \mathbb{Q}$. Plus précisément, il existe $v > 0$ tel que*

$$\mathbf{Var}[\omega_n] = vG_2(n) + O(1),$$

où G_2 est une fonction de propriétés différentes suivant la rationalité de $\log p/\log q$, de manière similaire à la fonction G_1 définie pour l'espérance.

Démonstration. On pose

$$R(z) = p_{\mathtt{m}}\frac{z^m}{m!}e^{-z},$$

si bien que

$$T(z,u) = T(pz,u)\,T(qz,u) + (u-1)R(z).$$

En dérivant 9.1 par rapport à u successivement une fois et deux fois, et en faisant $u = 1$, on obtient

$$\begin{aligned}
T_u(z,u) &= T_u(pz,u)\,T(qz,u) + T(pz,u)\,T_u(qz,u) + R(z)\\
T_u(z,1) &= T_u(pz,u)\,T(qz,u) + T(pz,u)\,T_u(qz,u) + R(z)\\
T_{u^2}(z,u) &= T_{u^2}(pz,u)\,T(qz,u) + T(pz,u)\,TT_{u^2}(qz,u) + 2T_u(pz,u)\,T_u(qz,u)\\
T_{u^2}(z,1) &= T_{u^2}(pz,u) + T_{u^2}(qz,u) + 2X(pz)X(qz)
\end{aligned}$$

$$\begin{aligned}
V(z) &= T_{u^2}(z,1) + T_u(z,1) - T_u^2(z,1)\\
&= T_{u^2}(z,1) + T_{u^2}(pz,u) + T_{u^2}(qz,u) + 2X(pz)X(qz) + X(z) - X(z)^2\\
&= V(pz) + V(qz) - X(pz) - X(qz) + X(pz)^2 + X(qz)^2 + 2X(pz)X(qz) + X(z) - X(z)^2\\
&= V(pz) + V(qz) + (X(pz) + X(qz))^2 - (X(pz) + X(qz)) + X(z) - X(z)^2\\
&= V(pz) + V(qz) - X(pz) - X(qz) + X(pz)^2 + X(qz)^2 + 2X(pz)X(qz) + X(z) - X(z)^2\\
&= V(pz) + V(qz) + (X(z) - R)^2 - (X(z) - R) + X(z) - X(z)^2\\
&= V(pz) + V(qz) - 2X(z)R(z) + R(z)^2 + R(z)
\end{aligned}$$

À l'aide du lemme 9.4, on sait que $V(z) \underset{|z|\to\infty}{=} O\left(|z|^{1+\varepsilon}\right)$ pour tout $\varepsilon > 0$. On a clairement $V(x) \underset{x\to 0}{=} O(x^2)$, donc la transformée de Mellin V^* de V admet $\langle -2, -1\rangle$ comme bande fondamentale.

Quelle est la nature de ses résidus ? On a vu, si $\mu = \frac{p_{\mathtt{m}}}{m(m-1)\left(\frac{1}{p}\log p + \frac{1}{q}\log q\right)}$, que

$$X(x) \underset{x\to\infty}{=} O(x^{-d}) + \mu\varpi_1(\log_2 x)\,x,$$

donc

$$X(x)R(x) \underset{x\to\infty}{=} p_{\mathrm{m}}\frac{x^m}{m!}e^{-x}O(x^{-d}) + p_{\mathrm{m}}\frac{x^m}{m!}e^{-x}\mu\varpi_1(\log_2 x)\,x,$$

et d'après (9.2), $\left(p_{\mathrm{m}}\frac{x^m}{m!}e^{-x}O(x^{-d}) + p_{\mathrm{m}}\frac{x^m}{m!}e^{-x}\mu\varpi_1(\log_2 x)\,x\right)^*$ n'a pas de résidus en $s=-1$.

D'autre part,

$$R(z) = p_{\mathrm{m}}\frac{z^m}{m!}e^{-z}, \quad R^2(z) = p_{\mathrm{m}}^2\frac{z^{2m}}{(m!)^2}e^{-2z},$$

donc

$$R^*(s) = \frac{p_{\mathrm{m}}}{m!}\Gamma(m+s), \quad (R^2)^*(s) = 2^{-(2m+s)}\left(\frac{p_{\mathrm{m}}}{m!}\right)^2\Gamma(2m+s).$$

Donc d'après (9.2), la transformée de Mellin

$$V^*(s) = \frac{R^*(s) + (R^2)^*(s) - 2(XR)^*(s)}{1-p^{-s}-q^{-s}}$$

n'a que des résidus simples, et la variance $V(x)$ est linéaire.

En dépoissonisant, on obtient que la variance dans le modèle de Bernoulli vérifie asymptotiquement

$$\mathbf{Var}[\omega_n] = V(n) - n(X'(n))^2 + O(1),$$

et est bien linéaire. □

9.2.2 Preuve résumée de la limite gaussienne par analyse complexe

On peut procéder exactement comme dans [78, 80, 113, 114, 146] pour montrer qu'on a une limite gaussienne.

Notons qu'en fait on a seulement besoin de l'asymptotique des moyenne et variance dans le modèle de Poisson. La dépoissonisation se fera en cours de notre analyse complexe multivariée.

On travaille alors directement sur $T(z,u)$, u restant dans un voisinage $V(1)$ de 1, pour montrer à l'aide de cônes croissants

$$\mathcal{C}_m = \left\{ z\in\mathcal{S}_\theta,\ A \leqslant \mathrm{Re}(z) \leqslant \frac{A}{pq^m}\right\} \times V(1),$$

où A un réel bien choisi, qu'on se trouve dans de bonnes conditions de dépoissonisation du type

$$z\in\mathcal{S}_\theta \implies |T(z,u)| > c_1|e^{-\alpha z}|, \tag{9.3}$$

$$z\notin\mathcal{S}_\theta \implies |T(z,u)e^{-z}| < c_2|e^{-\alpha z}|, \tag{9.4}$$

avec $c_1, c_2 > 0$, $\theta\in[0,\frac{\pi}{2}[$ et $\alpha\in]0,1[$.

On travaille ensuite sur $L(z,u) = \log T(z,u)$ pour montrer, toujours à l'aide de cônes croissants

$$\mathcal{C}_m = \left\{ (z,u)\ ;\ u\in V(1), z\in\mathcal{S}_\theta,\ \mathrm{Re}(z)\leqslant\frac{1}{q^m}\right\},$$

que la fonction $\frac{L(z,u)-\log u}{z}$ est bornée sur $\mathcal{S}_\theta\times V(1)$, où $V(1)$ est un voisinage de 1.

En écrivant, pour $u\in V(1)$ et $t\in V(0)$,

$$L(z,u) = iX(z)t - \frac{v(z)}{2}t^2 + g_z(t)t^3,$$

et

$$X(z) = \frac{1}{2i\pi} \oint_C \frac{L(z, e^{iw})}{w^2}\, dw$$

$$\frac{v(z)}{2} = \frac{1}{2i\pi} \oint_C \frac{L(z, e^{iw})}{w^3}\, dw$$

$$g_z(t) = \frac{1}{2i\pi} \oint_C \frac{L(z, e^{iw})}{w^3(w-t)}\, dw$$

où C est un contour inclus dans $V(0)$, on montre

$$X(z) = O(z) \text{ et } V(z) = O(z) \text{ pour } z \to \infty,\ z \in \mathcal{S}_\theta,$$

ce qui avec

$$X(x) = \underset{x\to 0}{O}(x^2)\ , \quad V(x) = \underset{x\to 0}{O}(x^2),$$

permet d'affirmer que les transformées de Mellin de X et V sont définies dans la bande $]-2, -1[$. On montre alors (voir ci-dessous la formule (9.2)) les asymptotiques

$$\begin{aligned} X(z) &\sim z\, Q_1(z), \\ V(z) &\sim z\, Q_2(z), \end{aligned}$$

avec Q_1 et Q_2 chacun bornés par des constantes positives.

On montre alors qu'on a une limite gaussienne pour la variable aléatoire ω dans le modèle de Poisson en passant par la fonction caractéristique, autrement dit

$$T(z, e^{it/\sqrt{v(z)}}) e^{-itX(z)} \underset{z\to\infty}{\to} e^{-\frac{t^2}{2}}.$$

On revient au modèle de Bernoulli en évaluant

$$T_n(u) = \frac{n!}{2i\pi} \oint_{|z|=n} T(z, u) e^z\, \frac{dz}{z^{n+1}}.$$

Pour ce faire, on écrit

$$\begin{cases} X(ne^{i\theta}) &= X(n) + nX'(n)i\theta + O(n\theta^2), \\ V(ne^{i\theta}) &= V(n) + O(n\theta), \end{cases}$$

dans un cône $\mathcal{S}_{\theta_{\max}}$, et on utilise les inégalités (9.3), qui permettent de négliger ce qui se passe en dehors du cône. On montre alors la convergence de la fonction caractéristique vers celle d'une loi normale avec certains paramètres. On déduit de ces paramètres que la moyenne $X_B(n)$ et la variance $V_B(n)$ de notre variable aléatoire dans le modèle de Bernouilli satisfont

$$\begin{cases} X_B(n) &\sim X(n), \\ V_B(n) &\sim V(n) - nX'(n)^2, \\ \alpha n &\leqslant V_B(n) \leqslant \beta n, \end{cases}$$

pour certaines constantes α et β strictement positives.

On a ainsi effectué une dépoissonisation au passage, mais en montrant au préalable la convergence vers une limite normale.

9.2.3 Preuve de la limite gaussienne par méthode de point fixe

Nous montrons que le nombre d'occurrences d'un motif dans les tries dans le modèle de Bernoulli suit une limite gaussienne, en utilisant cette fois la méthode de point fixe. C'est pour cette méthode que nous avons eu besoin d'une expansion asymptotique des grandeurs dépoissonisées, et d'établir des conditions de dépoissonisation plus fortes pour $X(z)$ et $V(z) + X^2(z)$.

Théorème 9.8 *ω_n possède une limite gaussienne.*

Démonstration. On peut procéder *mutatis mutandis* comme dans [114], où est établie la limite gaussienne de la taille des tries à l'aide d'une analyse complexe bivariée. Pour fixer les idées par rapport à la prochaine section, nous allons donner une autre preuve en utilisant des méthodes de point fixe, en procédant comme dans [169], qui ne traitait que le cas $p = q = 1/2$, en généralisant au cas $\log p/\log q$ rationnel. Notons que b_n est nul à partir du rang $n = m + 1$, et qu'on peut donc en faire abstraction.

ω_n vérifie la récurrence

$$\omega_n \stackrel{\mathcal{D}}{=} \omega^{(1)}_{I_1^{(n)}} + \omega^{(2)}_{I_2^{(n)}} + b_n,$$

où b_n est une variable aléatoire valant 1 si le trie considéré à n clés est égal à l'arbre $\mathfrak{m}$, et vaut zéro sinon. Selon les notations de (B.1),

$$\mathbf{E}\omega_n = f(n) + o(g^{1/2}(n)),\ \mathbf{Var}(\omega_n) = g(n) + o(g(n)),$$

avec $f(n) = n\omega_1(\log_\gamma n)$ et $g(n) = n\omega_2(\log_\gamma n)$.

On veut montrer qu'il existe $s \in]2, 3]$ tel que

$$\frac{1}{g^{1/2}(n)}\left(b_n - f(n) + \sum_{r=1}^{K} f(I_r^{(n)})\right) \xrightarrow{L_s} 0 \tag{9.5}$$

et

$$\left(\frac{g(I_r^{(n)})}{g(n)}\right) \xrightarrow{L_s} A_r^*, \tag{9.6}$$

avec

$$\sum_{r=1}^{K}(A_r^*)^2 = 1,\ \mathbf{Pr}(\exists r,\ A_r^* = 1) < 1. \tag{9.7}$$

Montrons (9.5) avec $s = 3$. On distinguera les cas $\log p/\log q$ rationnel et $\log p/\log q$ irrationnel.

Cas $\log p/\log q$ rationnel. —

Écrivons

$$G_1(n) = \mu_1 n\varpi_1(\log_\gamma n) + \sum_{k=2}^{M} \mu_k n^{r_k}\varpi_k(\log_\gamma n),$$

où les r_k sont définies par $\log x_k$ où les x_k sont les racines du polynôme $X^{s_1} + X^{s_2} - 1$.

Montrons d'abord

$$I_1\varpi_1(\log_\gamma I_1) + I_2\varpi_1(\log_\gamma I_2) - n\varpi_1(\log_\gamma n) \xrightarrow{L_s} 0.$$

Soient

$$E_n = \left\{ \left| I_1^{(n)} - np \right| < n^{3/4} \right\} = \left\{ \left| I_2^{(n)} - nq \right| < n^{3/4} \right\}$$

$$F_n = \left\{ \left| I_1^{(n)} - np \right| \geqslant n^{3/4} \right\} = \left\{ \left| I_2^{(n)} - nq \right| \geqslant n^{3/4} \right\}$$

$$I_1^{(n)} \sim \mathcal{B}(n,p)$$

$$I_2^{(n)} = n - I_1^{(n)} \sim \mathcal{B}(n,q)$$

On pose $\pi_1 = p$, $\pi_2 = q$.

$$a_r = I_r^{(n)}, \quad b_r = n\pi_r.$$

$$\eta(x) = \pi_r + \log_\gamma \frac{x}{n}, \ \eta'(x) = \frac{1}{x \log \gamma}.$$

Il existe c_r entre a_r et b_r

$$\begin{aligned} \eta'(c_r) &= \frac{\eta(b_r) - \eta(a_r)}{b_r - a_r} \\ &= \frac{\pi_r + \log_\gamma \pi_r - \pi_r - \log_\gamma I_r^{(n)}}{n\pi_r - I_r^{(n)}} = \frac{s_r + \log_\gamma I_r^{(n)}}{I_r^{(n)} - n\pi_r} \end{aligned}$$

d'où

$$s_r + \log_\gamma I_r^{(n)} = n\eta'(c_r) \frac{I_r^{(n)} - n\pi_r}{n} = \Psi_{nr} \frac{I_r^{(n)} - n\pi_r}{n}$$

avec

$$\Psi_{nr} = n\eta'(c_r) = \frac{n}{c_r \log \gamma}$$

Or c_r est entre $I_r^{(n)}$ et $n\pi_r$, et sur E_n on a

$$\left| I_r^{(n)} - n\pi_r \right| < n^{3/4},$$

donc à plus forte raison $\left| c_r - n\pi_r \right| < n^{3/4}$, soit

$$\begin{aligned} & -n^{3/4} < c_r - n\pi_r < n^{3/4} \\ \Leftrightarrow \ & -n^{3/4} + n\pi_r < c_r < n^{3/4} + n\pi_r \\ \Leftrightarrow \ & \frac{1}{n^{3/4} + n\pi_r} < \frac{1}{c_r} < \frac{1}{-n^{3/4} + n\pi_r} \\ \Leftrightarrow \ & \frac{n}{n^{3/4} + n\pi_r} < \frac{n}{c_r} < \frac{n}{-n^{3/4} + n\pi_r} \\ \Leftrightarrow \ & \frac{n}{n^{3/4} + n\pi_r} < \frac{n}{c_r} < \frac{n}{-n^{3/4} + n\pi_r} \\ \Leftrightarrow \ & \frac{1}{\pi_r} - \frac{n^{3/4}/\pi_r}{n^{3/4} + n\pi_r} < \frac{n}{c_r} < \frac{1}{\pi_r} + \frac{n^{3/4}/\pi_r}{n^{3/4} + n\pi_r}. \end{aligned}$$

On a donc, sur E_n,

$$\left\| \Psi_{nr} - \frac{1}{\pi_r \log \gamma} \right\|_\infty < \frac{1}{\log \gamma} \frac{n^{3/4}/\pi_r}{n^{3/4} + n\pi_r} < \pi_r^{-2} n^{-1/4} \log^{-1} \gamma.$$

Il s'ensuit que, pour n suffisamment grand, $|\Psi_{nr}|$ est borné par une constante M. Donc, toujours sur E_n,

$$\left|s_r + \log_\gamma I_r^{(n)}\right| = \left|\Psi_{nr}\right| \left|\frac{I_r^{(n)} - n\pi_r}{n}\right| < Mn^{-1/4},$$

ce qui implique, ω_1 étant 1-périodique,

$$\begin{aligned}
\omega_1\left(\log_\gamma I_r^{(n)}\right) &= \omega_1\left(s_r + \log_\gamma I_r^{(n)}\right) \\
&= \omega_1\left(\log_\gamma n + s_r + \log_\gamma \frac{I_r^{(n)}}{n}\right) \\
&= \omega_1\left(\log_\gamma n\right) + \Phi_{nr}\left(s_r + \log_\gamma \frac{I_r^{(n)}}{n}\right),
\end{aligned}$$

avec

$$\Phi_{nr} \in \left\{\omega_1'\left(x + \log_\gamma n\right) \mid |x| \leqslant Mn^{-1/4}\right\} =: B_n, \ r = 1, 2.$$

La continuité de ω_1' entraîne que les B_n sont des intervalles. Or, comme sa primitive, ω_1' est périodique donc uniformément continue, et la longueur des intervalles B_n tend vers zéro. Finalement, sur E_n, comme Φ_{n1} et Φ_{n2} appartiennent tous deux à B_n,

$$\|\Phi_{n1} - \Phi_{n2}\|_\infty \underset{n\to\infty}{\longrightarrow} 0.$$

Posons $d_{nr} = \Psi_{nr} - \frac{1}{\pi_r \log \gamma}$.

$$\begin{aligned}
& I_1\omega_1\left(\log_\gamma I_1\right) + I_2\omega_1\left(\log_\gamma I_2\right) - n\omega_1\left(\log_\gamma n\right) \\
= \ & I_1\omega_1\left(\log_\gamma n\right) + I_1\Phi_{n1}\left(s_1 + \log_\gamma \frac{I_1}{n}\right) + I_2\omega_1\left(\log_\gamma n\right) + I_2\Phi_{n1}\left(s_2 + \log_\gamma \frac{I_2}{n}\right) - n\omega_1\left(\log_\gamma n\right) \\
= \ & I_1\Phi_{n1}\left(s_1 + \log_\gamma \frac{I_1}{n}\right) + I_2\Phi_{n2}\left(s_2 + \log_\gamma \frac{I_2}{n}\right) \\
= \ & \Phi_{n1}\Psi_{n1}I_1\frac{I_1 - np}{n} + \Phi_{n2}\Psi_{n2}I_2\frac{I_2 - nq}{n} \\
= \ & \Phi_{n1}\left(d_1 + \frac{1}{\pi_1 \log\gamma}\right) I_1\frac{I_1 - np}{n} + \Phi_{n2}\left(d_2 + \frac{1}{\pi_2 \log\gamma}\right) I_2\frac{I_2 - nq}{n} \\
= \ & \frac{\Phi_{n1}}{n\log\gamma}\left(\frac{I_1}{\pi_1}\frac{I_1 - np}{n} + \frac{I_2}{\pi_2}\frac{I_2 - nq}{n}\right) + \frac{\Phi_{n2} - \Phi_{n1}}{\log\gamma}\frac{I_2}{\pi_2}\frac{I_2 - nq}{n} \\
& + \Phi_{n1}d_1I_1\frac{I_1 - np}{n} + \Phi_{n2}d_2\frac{I_2 - nq}{n}
\end{aligned}$$

Posons $R_n = \frac{I_1}{n} - p = q - \frac{I_2}{n}$, qui vérifie $\|R_n\|_\infty \leqslant n^{-1/4}$ sur E_n. Il vient

$$\begin{aligned}
\frac{I_1}{\pi_1}\frac{I_1 - np}{n} + \frac{I_2}{\pi_2}\frac{I_2 - nq}{n} &= \left(\frac{I_1}{p} - \frac{I_2}{q}\right) R_n \\
&= \left(\frac{qI_1 - pI_2}{pq}\right) R_n \\
&= \left(\frac{q(I_1 + I_2) - (p+q)I_2}{pq}\right) R_n \\
&= \frac{qn - I_2}{pq} R_n
\end{aligned}$$

Donc

$$\int_{E_n} \left| \frac{qn - I_2}{pq\sqrt{n}} R_n \right|^3 \, dP \leqslant \|R_n\|_\infty \frac{1}{\sqrt{pq}} \mathbf{E}\left[|\mathcal{N}(0,1)|^3 \right] + O(1) \underset{n\to\infty}{\longrightarrow} 0.$$

D'autre part,

$$\frac{1}{\sqrt{n}} \frac{\Phi_{n2} - \Phi_{n1}}{\log \gamma} \frac{I_2}{\pi_2} \frac{I_2 - nq}{n} = \frac{1}{\pi_2 \log \gamma} (\Phi_{n2} - \Phi_{n1}) \frac{I_2}{n} \frac{I_2 - nq}{\sqrt{n}}$$

avec $0 \leqslant \frac{I_2}{n} \leqslant 1$, donc

$$\int_{E_n} \left| \frac{1}{\pi_2 \log \gamma} (\Phi_{n2} - \Phi_{n1}) \frac{I_2}{n} \frac{I_2 - nq}{\sqrt{n}} \right|^3 \, dP \underset{n\to\infty}{\longrightarrow} 0.$$

De même, puisque $d_{nr} \underset{n\to\infty}{\longrightarrow} 0$ pour $r = 1, 2$, et que les Φ_{nr} sont bornés par $\|\omega_1'\|_\infty$, on a aussi

$$\frac{1}{\sqrt{n}} \int_{E_n} \left| \Phi_{nr} d_r I_r \frac{I_r - n\pi_r}{n} \right|^3 \, dP \underset{n\to\infty}{\longrightarrow} 0.$$

Chacun des termes étudiés ci-dessus tend vers zéro dans l'espace $L_3^{E_n}$ des fonctions continues munies de la métrique

$$f \to \int_{E_n} |f|^3 \, dP,$$

donc leur somme

$$\frac{1}{\sqrt{n}} \left(I_1 \omega_1 \left(\log_\gamma I_1 \right) + I_2 \omega_1 \left(\log_\gamma I_2 \right) - n \omega_1 \left(\log_\gamma n \right) \right)$$

aussi.

Montrons maintenant

$$I_1^{r_k} \varpi_k(\log_\gamma I_1) + I_2^{r_k} \varpi_k(\log_\gamma I_2) - n^{r_k} \varpi_k(\log_\gamma n) \xrightarrow{L_s} 0$$

pour toute racine r_k de $1 - p^{-s} - q^{-s}$. On a, avec des notations similaires à précédemment, en remarquant $(np)^{-r^k} + (nq)^{-r^k} = n^{-r^k}$,

$$\begin{aligned}
& I_1^{r_k} \varpi_k(\log_\gamma I_1) + I_2^{r_k} \varpi_k(\log_\gamma I_2) - n^{r_k} \varpi_k(\log_\gamma n) \\
= \; & I_1^{r_k} \Phi_{n1}\left(s_1 + \log_\gamma \frac{I_1}{n} \right) + I_2^{r_k} \Phi_{n1}\left(s_2 + \log_\gamma \frac{I_2}{n} \right) \\
& + r_k c_1^{r_k - 1} \Phi_{n1}\left(s_1 + \log_\gamma \frac{I_1}{n} \right) + r_k c_2^{r_k - 1} \Phi_{n2}\left(s_2 + \log_\gamma \frac{I_2}{n} \right) \\
= \; & \Phi_{n1} \Psi_{n1} I_1^{r_k} \frac{I_1 - np}{n} + \Phi_{n2} \Psi_{n2} I_2^{r_k} \frac{I_2 - nq}{n} \\
& + r_k c_1^{r_k - 1} \Phi_{n1}\left(s_1 + \log_\gamma \frac{I_1}{n} \right) + r_k c_2^{r_k - 1} \Phi_{n2}\left(s_2 + \log_\gamma \frac{I_2}{n} \right),
\end{aligned}$$

avec c_k compris entre I_k et $n\pi_k$. On obtient facilement, comme précédemment,

$$\Phi_{n1} \Psi_{n1} I_1^{r_k} \frac{I_1 - np}{n} + \Phi_{n2} \Psi_{n2} I_2^{r_k} \frac{I_2 - nq}{n} \xrightarrow{L_s} 0$$

sur E_n.

Les Ψ_{nk} étant bornés, on montre aussi facilement

$$r_k c_1^{r_k-1}\Phi_{n1}\left(s_1+\log_\gamma\frac{I_1}{n}\right)+r_k c_2^{r_k-1}\Phi_{n2}\left(s_2+\log_\gamma\frac{I_2}{n}\right)\xrightarrow{L_s}0$$

sur E_n.

Les convergences sur F_n sont évidentes.

Avec $m_2:=\min_{x\in[0,1]}\omega_2(x)>0$, $\frac{1}{g^{1/2}(n)}$ reste bornée par $\frac{1}{\sqrt{m_2}\sqrt{n}}$, donc

$$\frac{1}{g^{1/2}(n)}\left(I_1\omega_1\left(\log_\gamma I_1\right)+I_2\omega_1\left(\log_\gamma I_2\right)-n\omega_1\left(\log_\gamma n\right)\right)$$

tend aussi vers zéro dans $L_3^{E_n}$.

Or d'après les bornes de Chernoff [103], si S_n est la somme de n variables de Bernoulli indépendantes de même moyenne p, alors pour tout $0\leqslant\epsilon\leqslant 1$,

$$P[S_n\geqslant(1+\epsilon)np]\leqslant e^{-\epsilon^2np/3},$$

$$P[S_n\leqslant(1-\epsilon)np]\leqslant e^{-\epsilon^2np/2}.$$

Donc

$$P[F_n]\leqslant e^{-\epsilon^2np/3}+e^{-\epsilon^2np/2}\leqslant 2e^{-\epsilon^2np/2},$$

avec $\epsilon=p^{-1}n^{-1/4}$, ce qui entraîne

$$P[F_n]\leqslant 2e^{-\sqrt{n}/(2p)}.$$

Avec $m_2:=\min_{x\in[0,1]}\omega_2(x)>0$, nous obtenons pour $r=1,2$,

$$\int_{F_n}\left|\frac{1}{g^{1/2}(n)}\left(1-f(n)+f\left(I_1^{(n)}\right)+f\left(I_2^{(n)}\right)\right)\right|^3\,dP\leqslant\left(\frac{2\left\|\omega_1\right\|_\infty}{m_2^{1/2}}\right)^3 n^{3/2}2e^{-\sqrt{n}/(2p)}\underset{n\to\infty}{\longrightarrow}0.$$

Finalement, en rassemblant avec les résultats en-dessus,

$$\int\left|\frac{1}{g^{1/2}(n)}\left(1-f(n)+f\left(I_1^{(n)}\right)+f\left(I_2^{(n)}\right)\right)\right|^3\,dP\underset{n\to\infty}{\longrightarrow}0.$$

Montrons maintenant (9.6). Remarquons d'abord

$$\frac{g\left(I_r^{(n)}\right)}{g(n)}=\frac{I_r^{(n)}}{n}\frac{\omega_2\left(\log_\gamma\left(I_r^{(n)}\right)\right)}{\omega_2\left(\log_\gamma n\right)}\left(1-O(I_r^{a-1})\right),$$

où a est la plus petite partie réelle différente de -1, parmi les parties réelles des racines de $1-p^{-s}-q^{-s}$. (Donc $a-1<0$).

Par la loi simple des grands nombres et par convergence dominée, on a $\frac{I_r^{(n)}}{n}\to\pi_r$ dans tout L_p. En outre,

$$\begin{aligned}\left|\frac{\omega_2\left(\log_\gamma\left(I_r^{(n)}\right)\right)}{\omega_2\left(\log_\gamma n\right)}-1\right| &= \left|\frac{\omega_2\left(\log_\gamma n+s_r+\log_\gamma\left(\frac{I_r^{(n)}}{n}\right)\right)-\omega_2\left(\log_\gamma n\right)}{\omega_2\left(\log_\gamma n\right)}\right| \\ &= \frac{1}{m_2}\left|\omega_2\left(\log_\gamma n+s_r+\log_\gamma\left(\frac{I_r^{(n)}}{n}\right)\right)-\omega_2\left(\log_\gamma n\right)\right|\end{aligned}$$

Par la loi simple des grands nombres et la continuité uniforme de ω_2, cela tend presque sûrement vers zéro. Par convergence dominée, cette convergence reste vraie dans tout L_p.

En utilisant d'une part le fait que

$$\frac{\omega_2\big(\log_\gamma(I_r^{(n)})\big)}{\omega_2\left(\log_\gamma n\right)}$$

est borné et d'autre part

$$O(I_r^{a-1}) \xrightarrow{L_s} 0$$

en utilisant les bornes de Chernoff, on montre facilement que

$$\frac{\omega_2\big(\log_\gamma(I_r^{(n)})\big)}{\omega_2\left(\log_\gamma n\right)} \cdot O(I_r^{a-1})$$

converge vers 0 dans L_s.

Finalement, cela implique pour tout $s \in]2,3]$ (et donc pour $s=3$)

$$\frac{g\big(I_r^{(n)}\big)}{g(n)} \xrightarrow[n\to\infty]{L_s} \pi_r.$$

(9.7) est évident :

$$(\sqrt{\pi_1})^2 + (\sqrt{\pi_2})^2 = 1 \text{ et } \mathbf{Pr}(\exists r,\ \sqrt{\pi_r} = 1) < 1.$$

Cas $\log p/\log q$ irrationnel. —

La fonction f s'écrit

$$f(x) = \sum_{k\in\mathbb{Z}^*} \frac{p_{\mathfrak{m}}}{m!} \frac{\Gamma(m+s_k)}{h(s_k)} x^{-s_k}$$

où

$$h(t) = -p^{-t}\log p - q^{-t}\log q$$

et s_k sont les racines $\neq 1$ de $1 - p^{-s} - q^{-s}$ On a donc $f(np) + f(nq) - f(n) = 0$, et pour tout k il existe $c_1 \in [\min(k, np), \max(k, np)]$ et $c_2 \in [\min(n-k, nq), \max(n-k, nq)]$ tels que

$$\begin{aligned} f(k) + f(n-k) - f(n) &= f(k) - f(np) + f(n-k) - f(nq) \\ &= -\sum_{j\in\mathbb{Z}^*} s_j \frac{p_{\mathfrak{m}}}{m!} \frac{\Gamma(m+s_j)}{h(s_j)} c_1^{-s_j-1}(k-np) \\ &\quad -\sum_{j\in\mathbb{Z}^*} s_j \frac{p_{\mathfrak{m}}}{m!} \frac{\Gamma(m+s_j)}{h(s_j)} c_2^{-s_j-1}(nq-k) \end{aligned}$$

Il est clair que

$$\left|\frac{f(0) + f(n) - f(n)}{\sqrt{n}}\right|^3 \mathbf{Pr}[k=0] \longrightarrow 0$$

si n tend vers l'infini. D'autre part, de $\mathrm{Re}(s_j) > -1$, on déduit $\left|c_1^{-s_j-1}\right| < 1$ et $\left|c_2^{-s_j-1}\right| < 1$ dès que $k \neq 0$ et $k \neq n$. Montrons maintenant

$$\frac{f(k) + f(n-k) - f(n)}{\sqrt{n}} \xrightarrow{L_3} 0.$$

Soit $\varepsilon > 0$. À l'aide de

$$\forall x \in \mathbb{R}, \quad |\Gamma(x+iy)| \underset{y\to\pm\infty}{\sim} \sqrt{2\pi}|y|^{x-\frac{1}{2}}e^{-\pi|y|/2},$$

et toujours du lemme 9.1, qui entraîne notamment

$$|s_k|^2 < \max(1,\sigma_0)^2 + \left(\frac{(2k+1)\pi}{\log p}\right)^2,$$

on trouve $J \in \mathbb{N}^*$ tel que

$$\sum_{|j|>J} |s_j| \frac{p_m}{m!} \left|\frac{\Gamma(m+s_j)}{h(s_j)}\right| D < \varepsilon.$$

En outre, $\int \left|\frac{k-np}{\sqrt{n}}\right|^3 \to \mathbf{E}\sqrt{pq}\,|\mathcal{N}(0,1)|^3$, donc $\int \left|\frac{k-np}{\sqrt{n}}\right|^3$ est borné par une constante C. Donc, nous avons

$$\int_{k\notin\{0,n\}} \left|\sum_{|j|>J} s_j \frac{p_m}{m!} \frac{\Gamma(m+s_j)}{h(s_j)} c_1^{-s_j-1} \frac{k-np}{\sqrt{n}}\right|^3 d\mathbf{Pr} < C\varepsilon^3,$$

pour tout n.

Soit N tel que, pour $n > N$, on ait

$$\left|\frac{f(0)+f(n)-f(n)}{\sqrt{n}}\right|^3 \mathbf{Pr}[k=0] < \varepsilon.$$

Il existe $\alpha > 0$ tel que, pour tout j tel que $|j| \leqslant J$, $\mathrm{Re}(s_j) > -1+\alpha$. À partir de là, il est facile de montrer, en utilisant les bornes de Chernoff comme précédemment, que

$$\sum_{|j|>J} |s_j| \frac{p_m}{m!} \left|\frac{\Gamma(m+s_j)}{h(s_j)}\right| \frac{1}{|c_1|^{s_j+1}} \frac{k-np}{\sqrt{n}}$$

tend vers zéro dans L_3 lorsque n tend vers l'infini, et qu'il existe $N' > N$ tel que, pour $n > N'$,

$$\int \sum_{|j|\leqslant J} |s_j| \frac{p_m}{m!} \frac{1}{c_1^{s_j+1}} \frac{\Gamma(m+s_j)}{h(s_j)} \frac{k-np}{\sqrt{n}}\, d\mathbf{Pr} < \varepsilon.$$

Pour n assez grand, on a donc

$$\int \left|\frac{f(k)-f(np)}{\sqrt{n}}\right|^3 d\mathbf{Pr}$$

aussi petit qu'on veut. C'est aussi vrai pour

$$\int \left|\frac{f(n-k)-f(nq)}{\sqrt{n}}\right|^3 d\mathbf{Pr}.$$

On en déduit

$$\frac{f(k)+f(n-k)-f(n)}{\sqrt{n}} \xrightarrow{L_3} 0.$$

Les propriétés (9.6) et (9.7) sont évidentes car on a immédiatement

$$\frac{g(I_r^{(n)})}{g(n)} \xrightarrow[n\to\infty]{L_3} \pi_r.$$

□

9.3 Longueur de cheminement interne

On s'intéresse à la somme des hauteurs des nœuds internes dans un trie, correspondant à la somme des longueurs des préfixes stricts communs des mots aux feuilles. On considère une source de Bernoulli $\mathcal{B}(n,p)$, et on note $q = 1-p$.

Nous montrons la limite gaussienne de ce paramètre en suivant le même cheminement que dans la section précédente, si ce n'est que l'étude de la longueur de cheminement interne passe par l'étude de la taille des tries. Ainsi, on considère d'abord une source suivant une loi de Poisson. En utilisant une série génératrice multivariée, on trouve l'asymptotique de la matrice de covariance des taille et longueur de cheminement interne, d'abord dans le cas poissonien, puis dans le cas d'une source de Bernoulli, puis on utilise un théorème de point fixe multivarié.

En plus de la difficulté que rajoute l'étude d'un paramètre multivarié, apparaîtra une autre difficulté : la matrice de covariance ne sera pas linéaire mais légèrement surlinéaire, avec l'apparition de termes en $n\log n$ et $n\log^2 n$, rendant les calculs plus compliqués.

Nous donnons d'abord un lemme permettant de montrer l'analyticité d'une série entière.

Lemme 9.5 *Soient G et t deux séries entières telles que t possède un rayon de convergence infini, et*

$$G(z) = G(pz) + G(qz) + t(z).$$

Alors G possède aussi un rayon de convergence infini.

Démonstration. Écrivons $G(z) = \sum_{n\geqslant 0} g_n z^n$ et $t(z) = \sum_{n\geqslant 0} t_n z^n$. Pour tout $n \geqslant 2$ nous avons

$$g_n = \frac{t_n}{1-p^n-q^n}.$$

Or

$$\lim_{n\to\infty} \frac{1}{1-p^n-q^n} = 1,$$

donc

$$|g_n| < 2\,|t_n|$$

à partir d'un certain rang, et G possède aussi un rayon de convergence infini. □

Nous établissons maintenant des équations fonctionnelles sur les séries en z des moyennes, variances et covariance des paramètres qui nous intéressent, dans le cas d'une source de Poisson de paramètre z.

Théorème 9.9 *Plaçons-nous dans le cas poissonien tiré de la source de Bernoulli $\mathcal{B}(n,p)$, et soient I et L les taille et longueur de cheminement des tries. On définit*

$$\begin{aligned} e_2(z) &= 1 + z + \cdots + \frac{z^2}{2!} \\ f_2(z) &= e_2(z)e^{-z} \end{aligned}$$

et $R(z) = 1 - f_2(z)$. *Les séries* $X(z)$, $Y(z)$, $C(z)$ *et* $W(z)$ *représentant* $E[I]$, $E[L]$, $Var(I)$, $Cov(I,L)$ *et* $Var(L)$ *vérifient*

$$\begin{aligned}
X(z) &= X(pz) + X(qz) + 1 - e_2(z)e^{-z} = X(pz) + X(qz) + R(z), && (9.8)\\
Y(z) &= X(pz) + X(qz) + Y(pz) + Y(qz) = X(z) - R(z) + Y(pz) + Y(qz), && (9.9)\\
V(z) &= V(pz) + V(qz) + H(z), && (9.10)\\
C(z) &= C(pz) + C(qz) + V(z) - H(z) + (1 - R(z))Y(z), && (9.11)\\
W(z) &= W(pz) + W(qz) + 2C(z) - V(z) - (3 + 2R)Y(z) + H(z), && (9.12)
\end{aligned}$$

avec $H(z) = (2X - R)(1 - R)(z)$, *et sont analytiques sur* $\mathbb{C}$ *tout entier.*

Démonstration. Si T est un trie de sous-arbres T_1 et T_2, alors sa taille (nombre de nœuds internes) et sa longueur de cheminement interne vérifient

$$\begin{aligned}
I(T) &= I(T_0) + I(T_1) + 1 - \chi(|T| \geqslant b),\\
L(T) &= I(T_0) + I(T_1) + L(T_0) + L(T_1).
\end{aligned}$$

On en déduit facilement que la série génératrice bivariée des tries, où z est le paramètre de la source poissonienne, u dénombre la taille et v la longueur de cheminement interne,

$$P(z,u,v) = uP(pz,uv,v)P(qz,uv,v) + (1-u)e_2(z)e^{-z}.$$

En dérivant P par rapport respectivement à u et à v, on obtient

$$\begin{aligned}
P_u &= P^2 + uvP_uP + uvPP_u - e_2(z)e^{-z},\\
P_v &= u(uP_u + P_v)P + uP(uP_u + P_v).
\end{aligned}$$

En faisant $u = 1$ on trouve

$$\begin{aligned}
X(z) &= X(pz) + X(qz) + 1 - e_2(z)e^{-z} = X(pz) + X(qz) + R(z),\\
Y(z) &= X(pz) + X(qz) + Y(pz) + Y(qz) = X(z) - R(z) + Y(pz) + Y(qz).
\end{aligned}$$

En poursuivant les dérivations de P par rapport à u et v, on obtient

$$\begin{aligned}
P_{u^2} &= 2(vP_uP + vPP_u) + uv^2(P_{u^2}P + 2P_uP_u + PP_{u^2}),\\
P_{uv} &= (P_u + uP_{u^2} + P_{uv})P + (uP_u + P_v)P_u\\
&\quad + P_u(uP_u + P_v) + P(P_u + uP_{u^2} + P_{uv}) + (uP_u + P_v)P + P(uP_u + P_v),\\
P_{v^2} &= u(u(uP_{u^2} + P_{uv}) + uP_{uv} + P_{v^2})P + 2u(uP_u + P_v)(uP_u + P_v)\\
&\quad + uP(u(uP_{u^2} + P_{uv}) + uP_{uv} + P_{v^2}).
\end{aligned}$$

En $u = v = 1$, il vient

$$\begin{aligned}
P_{u^2} &= 2(P_u(pz) + P_u(qz)) + P_{u^2}(pz) + P_{u^2}(qz) + 2P_u(pz)P_u(qz) + P_{u^2}(qz),\\
P_{uv} &= P_u(pz) + uP_{u^2}(pz) + P_{uv}(pz) + (uP_u(pz) + P_v(pz))P_u(qz)\\
&\quad + P_u(pz)(P_u(qz) + P_v(qz)) + P_u(qz) + P_{u^2}(qz) + P_{uv}(qz)\\
&\quad + uP_u(pz) + P_v(pz) + P_u(qz) + P_v(qz),\\
P_{v^2} &= P_{u^2}(pz) + 2P_{uv}(pz) + P_{v^2}(pz) + 2(P_u(pz) + P_v(pz))(P_u(qz) + P_v(qz))\\
&\quad + P_{u^2}(qz) + 2P_{uv}(qz) + P_{v^2}(qz),
\end{aligned}$$

d'où

$$\begin{aligned}
V(z) &= P_{u^2} + P_u - P_u^2 \\
&= P_{u^2}(pz) + P_u(pz) - P_u^2(pz) + P_{u^2}(qz) + P_u(qz) - P_u^2(qz) \\
&\quad + P_u(pz) + P_u(qz) + 2P_u(pz)P_u(qz) + P_u - P_u^2 + P_u^2(pz) + P_u^2(qz) \\
&= V(pz) + V(qz) + X(pz) + X(qz) + 2X(pz)X(qz) + X(z) - X^2(pz) - X^2(qz) \\
&\quad - 2X(pz)X(qz) - 2X(pz)R - 2X(qz)R - R^2 \\
&= V(pz) + V(qz) + 2X(z) - R - 2(X-R)R - R^2 \\
&= V(pz) + V(qz) + 2X(z) - R - 2XR + 2R^2 - R^2 \\
&= V(pz) + V(qz) + 2X(z)(R-1) + R(R-1) \\
V(z) &= V(pz) + V(qz) + (2X - R)(1-R),
\end{aligned}$$

$$\begin{aligned}
C(z) &= P_{uv} - P_u P_v \\
&= P_{uv}(pz) + P_{uv}(qz) - (P_u(pz) + P_u(qz) + R)(P_u(pz) + P_u(qz) + P_v(pz) + P_v(qz)) \\
&\quad + \cdots \\
&= C(pz) + C(qz) - P_u^2(pz) - P_u(pz)P_u(qz) - P_u(pz)P_v(qz) - P_u(qz)P_u(pz) \\
&\quad - P_u^2(qz) - P_u(qz)P_v(pz) - RP_v(z) + P_u(pz) + P_{u^2}(pz) \\
&\quad + (P_u(pz) + P_v(pz))P_u(qz) + P_u(pz)(P_u(qz) + P_v(qz)) + P_u(qz) + P_{u^2}(qz) \\
&\quad + (X+Y)(pz) + (X+Y)(qz) \\
&= C(pz) + C(qz) + V(pz) + V(qz) + Y(z) - RY(z) \\
C(z) &= C(pz) + C(qz) + V(z) - (2X(z) - R(z))(1 - R(z)) + (1-R)Y(z),
\end{aligned}$$

et

$$\begin{aligned}
W(z) &= P_{v^2} + P_v - P_v^2 \\
&= P_{u^2}(pz) + 2P_{uv}(pz) + P_{v^2}(pz) + P_{u^2}(qz) + 2P_{uv}(qz) + P_{v^2}(qz) + 2(P_u + P_v)(P_u + P_v) \\
&\quad + P_v - (X(pz) + X(qz) + Y(pz) + Y(qz))^2 \\
&= \cdots \\
&= W(pz) + W(qz) - X(pz) - X(qz) + V(pz) + V(qz) \\
&\quad + 2(C(pz) + C(qz)) - Y(pz) - Y(qz) \\
&= W(pz) + W(qz) - Y(z) + V(z) - (2X(z) - R(z))(1 - R(z)) \\
&\quad + 2(C(z) - V(z) + (2X(z) - R(z))(1 - R(z)) - Y(z) - RP_v(z)) \\
&= W(pz) + W(qz) - 3Y(z) - V(z) + (2X(z) - R(z))(1 - R(z)) \\
&\quad + 2C(z) - 2RP_v(z) \\
W(z) &= W(pz) + W(qz) + 2C(z) - V(z) - (3 + 2R)Y(z) + H(z).
\end{aligned}$$

Le lemme 9.5 permet de montrer successivement que les séries ci-dessus ont un rayon de convergence infini. □

À l'aide des équations trouvées précédemment, nous déterminons maintenant l'asymptotique des séries qui nous intéressent.

Théorème 9.10 *Il existe des constantes strictement positives* $m_0^{(P)}$, $m_1^{(P)}$, $m_{12}^{(P)}$, $v^{(P)}$, $c^{(P)}$, $w^{(P)}$, *et des fonctions* $G_X^{(P)}$, $G_Y^{(P)}$, $G_{Y2}^{(P)}$, $G_V^{(P)}$, $G_C^{(P)}$, $G_W^{(P)}$, *en* $O(z^{-\epsilon})$ *(pour* $z \to \infty$*) avec* $\epsilon > 0$

si $\log p/\log q \notin \mathbb{Q}$, *périodiques en* $\log_\gamma(n)$ *d'amplitude négligeable devant 1 sinon, telles que*

$$X(z) = m_0^{(P)}\left(1+G_X^{(P)}(z)\right)z+O(1), \tag{9.13}$$

$$Y(z) = m_1^{(P)}\left(1+G_Y^{(P)}(z)\right)z\log(z)+m_{12}^{(P)}\left(1+G_{Y2}^{(P)}(z)\right)z+O(1), \tag{9.14}$$

$$V(z) = v^{(P)}\left(1+G_V^{(P)}(z)\right)z+O(1), \tag{9.15}$$

$$C(z) = c^{(P)}\left(1+G_C^{(P)}(z)\right)z\log(z)+O\left(\log(z)\right), \tag{9.16}$$

$$W(z) = w^{(P)}\left(1+G_W^{(P)}(z)\right)z\log^2(z)+O\left(z\log(z)\right). \tag{9.17}$$

En outre, si on note

$$h = p\log\frac{1}{p}+q\log\frac{1}{q},$$

on a les relations suivantes

$$m_1^{(P)} = \frac{m_0^{(P)}}{h}, \tag{9.18}$$

$$c^{(P)} = \frac{v^{(P)}}{h}, \tag{9.19}$$

$$w^{(P)} = \frac{v^{(P)}}{h^2}. \tag{9.20}$$

Remarque. Nous gardons les deux premiers termes dominants dans $Y(z)$ afin d'avoir un terme résiduel $O(1)$ négligeable devant la racine carrée de la variance.

Démonstration. Il est clair que les séries considérées sont en $O\left(|z|^{1+\varepsilon}\right)$ lorsque z tend vers ∞ dans un cône linéaire, grâce à une application successive du lemme 9.4, et qu'elles sont en $O\left(|z|^2\right)$ lorsque z tend vers 0.

Leurs transformées de Mellin respectives admettent donc toutes $\langle -2,-1\rangle$ comme bande fondamentale.

En prenant la transformée de Mellin de $X(z)$ et de son expression (9.8), on obtient

$$\begin{aligned} X^*(s) &= p^{-s}X^*(s)+q^{-s}X^*(s)-\Gamma(s)-\Gamma(s+1) \\ &= p^{-s}X^*(s)+q^{-s}X^*(s)-(s+1)\Gamma(s) \\ &= \frac{(s+1)\Gamma(s)}{p^{-s}+q^{-s}-1} = \frac{\Gamma(s+2)}{s\left(p^{-s}+q^{-s}-1\right)}. \end{aligned}$$

Soient $c\in]-2,-1[$, et s_k les racines de $p^{-s}+q^{-s}-1$. En inversant la transformation de Mellin $X^*(z)$, on en déduit l'expression de $X(z)$

$$\begin{aligned} X(z) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\Gamma(s+2)}{s\left(p^{-s}+q^{-s}-1\right)}z^{-s}\,ds \\ &= -\sum_{k\in\mathbb{Z}}\operatorname{Res}\frac{\Gamma(s+2)}{s\left(p^{-s}+q^{-s}-1\right)}z^{-s} \\ &= \sum_{k\in\mathbb{Z}}\frac{\Gamma(s_k+2)}{s_k h(s_k)}z^{-s_k}+O\left(z^{-d}\right), \end{aligned}$$

pour tout $d>0$.

De même, la transformée de Mellin de $Y(z)$ s'écrit

$$\begin{aligned} Y^*(s) &= X^*(s) + (s+1)\Gamma(s) + p^{-s}Y^*(s) + q^{-s}Y^*(s) \\ &= -\frac{X^*(s) + (s+1)\Gamma(s)}{p^{-s}+q^{-s}-1} \\ &= -\frac{(s+1)\Gamma(s)}{(p^{-s}+q^{-s}-1)^2} - \frac{(s+1)\Gamma(s)}{p^{-s}+q^{-s}-1}. \end{aligned}$$

On a des pôles doubles provenant de $\frac{1}{(p^{-s}+q^{-s}-1)^2}$, d'où un terme dominant en $z\log z$:

$$\begin{aligned} Y(z) &= \sum_{k\in\mathbb{Z}} \frac{\Gamma(s_k+2)}{s_k h^2(s_k)} z^{-s_k}\log z \\ &\quad + \sum_{k\in\mathbb{Z}} \left(\frac{\Gamma'(s_k+2)}{s_k h^2(s_k)} - \frac{h'(s_k)\Gamma(s_k+2)}{s_k h^2(s_k)} - \frac{\Gamma(s_k+2)}{s_k^2 h^2(s_k)} - \frac{\Gamma(s_k+2)}{s_k h(s_k)} \right) z^{-s_k} + O\left(z^{-d}\right). \end{aligned}$$

De plus, le fait d'avoir

$$Y^*(s) = -\frac{X^*(s)}{p^{-s}+q^{-s}-1} + \cdots$$

avec

$$h = -\left(p^{-s}+q^{-s}-1\right)'(-1),$$

entraîne

$$m_1^{(P)} = \frac{m_0^{(P)}}{h}.$$

En outre, d'après le lemme 9.1, dans le cas $\log p/\log q$ rationnel, les racines z_k de $s \mapsto p^{-s}+q^{-s}-1$ sont de la forme

$$z_k = -1 + \frac{2ks_1 i\pi}{\log p} = -1 + \frac{2kti\pi}{\log p}$$

avec k, s, t entiers, donc on a aussi

$$\forall k,\ -\left(p^{-s}+q^{-s}-1\right)'(z_k) = \log p p^{-z_k} + \log q q^{-z_k} = h, \tag{9.21}$$

d'où $G_Y^{(P)} = G_X^{(P)}$.

La transformée de Mellin de $Y(z)$ s'écrit

$$\begin{aligned} V^*(s) &= p^{-s}V^*(s) + q^{-s}V^*(s) + H^*(s) \\ &= -\frac{H^*(s)}{p^{-s}+q^{-s}-1} \end{aligned}$$

On en déduit l'expression de $V(z)$ par inversion de la transformation de Mellin. La transformée de Mellin de $C(z)$ s'écrit

$$\begin{aligned} C^*(s) &= p^{-s}C^*(s) + q^{-s}C^*(s) + V^*(s) - H^*(s) + \mathcal{M}((1-R)Y) \\ &= \frac{H*(s) - V^*(s) - \mathcal{M}((1-R)Y)}{p^{-s}+q^{-s}-1} \\ &= \frac{H^*(s)}{(p^{-s}+q^{-s}-1)^2} + \frac{H^*(s) - \mathcal{M}((1-R)Y)}{p^{-s}+q^{-s}-1}, \end{aligned}$$

et les pôles doubles de $\frac{1}{p^{-s}+q^{-s}-1}$ donnent un terme dominant en $z\log z$ pour $C(z)$, avec

$$c^{(P)} = \frac{v^{(P)}}{h},$$

suivant un raisonnement similaire au précédent.

$$\begin{aligned} W^*(s) &= p^{-s}W^*(s) + q^{-s}W^*(s) + 2C^*(s) - V^*(s) + \mathcal{M}((3+2R)Y) + H^*(s) \\ &= \frac{2C^*(s) - V^*(s) + \mathcal{M}((3+2R)Y) + H^*(s)}{1 - p^{-s} - q^{-s}} \\ &= \frac{2H^*(s)}{(1-p^{-s}-q^{-s})^3} + \cdots \end{aligned}$$

donc on a des pôles d'ordre 3 et un terme dominant en $z\log^2 z$. En utilisant la proposition 9.1, on a la relation

$$w^{(P)} = \frac{v^{(P)}}{h^2}.$$

□

9.3.1 Dépoissonisation et matrice de covariance

Nous pouvons donner l'asymptotique des coefficients la matrice de covariance de la distribution jointe de la taille et de la longueur de cheminement des tries, en dépoissonisant les asymptotiques trouvées plus haut.

Théorème 9.11 *Il existe des constantes strictement positives* $m_0^{(B)}$, $m_1^{(B)}$, $m_{12}^{(B)}$, $v^{(B)}$, $c^{(B)}$, $w^{(B)}$, *et des fonctions* $G_X^{(B)}$, $G_Y^{(B)}$, $G_{Y2}^{(B)}$, $G_V^{(B)}$, $G_C^{(B)}$, $G_W^{(B)}$, *en* $O(n^{-\epsilon})$ *avec* $\epsilon > 0$ *si* $\log p/\log q \notin \mathbb{Q}$, *périodiques en* $\log_\gamma(n)$ *d'amplitude négligeable devant 1 sinon, telles que*

$$\begin{aligned} X_B(n) &= m_0^{(B)}\left(1 + G_X^{(B)}(n)\right)n + O(1), \\ Y_B(n) &= m_1^{(B)}\left(1 + G_Y^{(B)}(n)\right)n\log(n) + m_{12}^{(B)}\left(1 + G_{Y2}^{(B)}(n)\right)n + O(1), \\ V_B(n) &= v^{(B)}\left(1 + G_V^{(B)}(n)\right)n + O(1), \\ C_B(n) &= c^{(B)}\left(1 + G_C^{(B)}(n)\right)n\log(n) + O\left(n\right), \\ W_B(n) &= w^{(B)}\left(1 + G_W^{(B)}(n)\right)n\log^2(n) + O\left(n\log(n)\right). \end{aligned}$$

En outre, on a $m_1^{(B)} = m_0^{(B)}/h$, $c^{(B)} = v^{(B)}/h$ *et* $w^{(B)} = v^{(B)}/h^2$.

Démonstration. À l'aide d'une application successive du lemme 9.3 et de la remarque 9.1, on montre que les séries considérées vérifients les conditions de dépoissonisation du théorème 9.1, et le cas échéant des théorèmes 9.2 et 9.3 pour les trois dernières séries.

Après dépoissonisation, on trouve

$$\begin{aligned} m_0^{(B)} &= m_0^{(B)}, \\ m_1^{(B)} &= m_1^{(B)}, \\ m_{12}^{(B)} &= m_{12}^{(B)}, \\ v^{(B)} &= v^{(B)} - m_0^{(B)^2}, \\ c^{(B)} &= c^{(B)} - m_0^{(B)}m_1^{(B)} = \frac{v^{(B)}}{h} - m_0^{(B)}\frac{m_0^{(B)}}{h} = \frac{v^{(B)}}{h}, \\ w^{(B)} &= w^{(B)} - m_1^{(B)^2} = \frac{v^{(B)}}{h^2} - \frac{m_0^{(B)^2}}{h^2} = \frac{v^{(B)}}{h^2}. \end{aligned}$$

On sait que $v^{(B)}$ est non nul par les travaux de Régnier, Jacquet [113, 114]. Les expressions $c^{(B)} = \frac{v^{(B)}}{h}$ et $w^{(B)} = \frac{v^{(B)}}{h^2}$ montrent alors $c^{(B)} > 0$ et $w^{(B)} > 0$. □

Il ne nous a pas été possible de généraliser la méthode présentée dans la sous-section 9.2.2. En effet la présence du terme uv dans la formule (9.13) nous empêche de définir des cônes croissants sympathiques : en itérant des cônes croissants, on n'a plus l'assurance que le terme uv reste dans un voisinage borné de 1. Il semble donc nécessaire d'utiliser des méthodes de point fixe.

Contrairement à ce qu'on obtenait avec des méthodes d'analyse complexe, avec les méthodes de point fixe, on ne sait rien *a priori* sur la loi limite du paramètre étudié dans le cas d'une source suivant une loi de Poisson

À l'aide des asymptotiques trouvées précédemment, nous pouvons maintenant montrer que la longueur de cheminement interne possède une loi limite gaussienne avec une source de Bernoullli. Nous aurons au cours de la preuve à distinguer les cas $\log p/\log q \in \mathbb{Q}$ et $\log p/\log q \notin \mathbb{Q}$.

Afin d'écourter la preuve du théorème ci-dessous, nous avons renvoyé la démonstration de certains résultats intermédaires dans des lemmes ci-après.

9.3.2 Une limite gaussienne

Théorème 9.12 *La loi jointe de la taille et de la longueur de cheminement interne d'un trie est asymptotiquement gaussienne.*

Démonstration. On note $L_{0,n}$ la variable aléatoire désignant la taille d'un trie, et $L_{1,n}$ la variable aléatoire désignant la longueur de cheminement interne d'un trie. Soit I_n la variable aléatoire égale aux nombre de clés du sous-arbre gauche. On a les relations

$$\begin{cases} L_{0,n} &= L_{0,I_n} + L_{0,n-I_n} + 1, \\ L_{1,n} &= L_{0,I_n} + L_{0,n-I_n} + L_{1,I_n} + L_{1,n-I_n}, \end{cases}$$

qui se traduisent matriciellement par :

$$\begin{pmatrix} L_{0,n} \\ L_{1,n} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} L_{0,I_n} \\ L_{1,I_n} \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} L_{0,n-I_n} \\ L_{1,n-I_n} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Notons

$$\begin{aligned} \mu_{0,n} &= m_0^{(B)} \left(1 + \varpi_X^{(B)}(\log_\gamma(n))\right), \\ \mu_{1,n} &= m_1^{(B)} \left(1 + \varpi_Y^{(B)}(\log_\gamma(n))\right), \\ \mu_{2,n} &= m_{12}^{(B)} \left(1 + \varpi_{Y2}^{(B)}(\log_\gamma(n))\right), \\ v_n &= v^{(B)} \left(1 + \varpi_V^{(B)}(\log_\gamma(n))\right), \\ c_n &= c^{(B)} \left(1 + \varpi_C^{(B)}(\log_\gamma(n))\right), \\ w_n &= w^{(B)} \left(1 + \varpi_W^{(B)}(\log_\gamma(n))\right). \end{aligned}$$

On pose

$$C_n = \begin{bmatrix} v_n n & c_n n \log n \\ c_n n \log n & w_n n \log^2 n \end{bmatrix}.$$

Pour alléger les notations, lorsque les fonctions périodiques n'interviennent pas de manière critique, on notera parfois

$$C_n = \begin{bmatrix} vn & cn \log n \\ cn \log n & wn \log^2 n \end{bmatrix}.$$

Comme on va normaliser par $C_n^{-1/2}$, on cherche la racine carrée S_n de C_n, de la forme $P\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2})P^{-1}$ si $C_n = P\mathrm{diag}(\lambda_1, \lambda_2)P^{-1}$. Pour calculer S_n, on procède comme dans [138], en utilisant le théorème de Cayley-Hamilton, d'après lequel S_n vérifie

$$S_n^2 - \mathrm{Tr}(S_n)S_n + \mathrm{dét}(S_n)\mathrm{Id}_2 = 0,$$

donc

$$S_n = \mathrm{Tr}(S_n)^{-1}(C_n + \mathrm{dét}(S_n)\mathrm{Id}_2),$$

avec $\mathrm{dét}(S_n) = \sqrt{\mathrm{dét}(C_n)}$, puisque $\mathrm{dét}(S_n) > 0$, et

$$\mathrm{Tr}(S_n^2) - \mathrm{Tr}(S_n)^2 + 2\mathrm{dét}(S_n) = 0,$$

soit $\mathrm{Tr}(S_n) = \sqrt{\mathrm{Tr}(C_n) + 2\mathrm{dét}(S_n)}$, puisque $\mathrm{Tr}(S_n) > 0$.

On pose $C_n = \begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix}$, et $d = vw - c^2$, d'où

$$\begin{aligned} S_n &= \frac{1}{\mathrm{Tr}(S_n)}\left(C_n + \mathrm{dét}(S_n)\mathrm{Id}_2\right) \\ &= \frac{1}{\sqrt{\alpha+\beta+2\sqrt{\alpha\beta-\gamma^2}}}\left(C_n + \sqrt{\alpha\beta-\gamma^2}\mathrm{Id}_2\right) \\ &= \frac{\sqrt{n}}{\sqrt{v\log n + w\log^2 n + 2\sqrt{d}\log n}} \begin{bmatrix} (v+\sqrt{d})\log n & c\log n \\ c\log n & w\log^2 n + \sqrt{d}\log n \end{bmatrix} \\ &= \frac{\sqrt{n\log n}}{\sqrt{v+2\sqrt{d}+w\log n}} \begin{bmatrix} v+\sqrt{d} & c \\ c & w\log n + \sqrt{d} \end{bmatrix}. \end{aligned}$$

Soient $M_n = \binom{\mu_0 n}{\mu_1 n\log n + \mu_1' n}$, $Y_n = \binom{L_{0n}}{L_{1n}}$, et $X_n = C_n^{-1/2}(Y_n - M_n)$. On pose

$$A_1 = A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Nous avons

$$\begin{aligned} X_n &= C_n^{-1/2}(Y_n - M_n) \\ &= A_1^{(n)} X_{I_1(n)} + A_2^{(n)} X_{I_2(n)} + b^{(n)}, \end{aligned}$$

avec

$$\begin{aligned} A_1^{(n)} &= C_n^{-1/2} A_1 C_{I_1}^{1/2}, \\ A_2^{(n)} &= C_n^{-1/2} A_2 C_{I_2}^{1/2}, \\ b^{(n)} &= A_1^{(n)} C_{I_1}^{-1/2} M_{I_1} + A_2^{(n)} C_{I_2}^{-1/2} M_{I_2} - C_n^{-1/2}\left(M_n + \binom{1}{0}\right). \end{aligned}$$

On refait les preuves comme précédemment.

Commençons par montrer les convergences du type

$$\frac{g(I_r^{(n)})}{g(n)} \underset{n\to\infty}{\longrightarrow} \pi_r.$$

Remarquons que si ϖ_1 est périodique, son carré l'est aussi, et que

$$\frac{\omega_2\left(\log_\gamma(I_r^{(n)})\right)}{\omega_2\left(\log_\gamma n\right)}$$

tend vers 1 dans tout L_p, donc on voit que les fonctions périodiques n'interviennent pas dans le calcul de la limite de $A_r^{(n)}$.

Avec cette remarque, et muni des lemmes 9.6 et 9.7 en-dessous, on montre facilement

$$A_r^{(n)} \longrightarrow diag(\sqrt{\pi}_r, \sqrt{\pi}_r)$$

en effectuant les calculs des matrices coefficient par coefficient. (Noter que $A_2^{(n)}$ s'obtient à partir de $A_1^{(n)}$ en remplaçant k par $n-k$).

En effet, on a

$$\begin{aligned}
&\left(\sqrt{\frac{k}{n}}\sqrt{\frac{v+2\sqrt{d}\log n+w\log^2 n}{v+2\sqrt{d}\log k+w\log^2 k}}\right)^{-1}(A_1^{(n)})_{11}\\
=\;&\Big(\frac{w\log(n)+\sqrt{d}}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
&-\frac{c}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\Big)\left(v+\sqrt{d}\log k\right)\\
&-\frac{c^2\log k}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
=\;&\frac{\sqrt{d}w\log k}{\sqrt{d}w\log n}+O\left(\frac{1}{\log n}\right)=\frac{\log k}{\log n}+O\left(\frac{1}{\log n}\right),
\end{aligned}$$

$$\begin{aligned}
&\left(\sqrt{\frac{k}{n}}\sqrt{\frac{v+2\sqrt{d}\log n+w\log^2 n}{v+2\sqrt{d}\log k+w\log^2 k}}\right)^{-1}(A_1^{(n)})_{12}\\
=\;&\Big(\frac{w\log(n)+\sqrt{d}}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
&-\frac{c}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\Big)c\log k\\
&-\frac{c\left(w\log^2 k+\sqrt{d}\log k\right)}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
=\;&\frac{c}{\sqrt{d}}\frac{\log k}{\log n}-\frac{c}{\sqrt{d}}\frac{\log^2 k}{\log^2 n}+O\left(\frac{1}{\log n}\right)\\
=\;&O\left(\frac{1}{\log n}\right),
\end{aligned}$$

$$
\begin{aligned}
&\left(\sqrt{\frac{k}{n}}\sqrt{\frac{v+2\sqrt{d}\log n+w\log^2 n}{v+2\sqrt{d}\log k+w\log^2 k}}\right)^{-1}(A_1^{(n)})_{21}\\
=\;&\Big(\frac{v+\sqrt{d}\log(n)}{\log n\left(vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)\right)}\\
&-\frac{c}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\Big)\left(v+\sqrt{d}\log k\right)\\
&+\frac{\left(v+\sqrt{d}\log(n)\right)c\log k}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
=\;&O\left(\frac{1}{\log n}\right),
\end{aligned}
$$

et

$$
\begin{aligned}
&\left(\sqrt{\frac{k}{n}}\sqrt{\frac{v+2\sqrt{d}\log n+w\log^2 n}{v+2\sqrt{d}\log k+w\log^2 k}}\right)^{-1}(A_1^{(n)})_{22}\\
=\;&\Big(\frac{v+\sqrt{d}\log(n)}{\log n\left(vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)\right)}\\
&-\frac{c}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\Big)c\log k\\
&+\frac{\left(v+\sqrt{d}\log(n)\right)\left(w\log^2 k+\sqrt{d}\log(k)\right)\log k}{vw\log(n)+v\sqrt{d}+(\log(n))^2\sqrt{d}w+\log(n)d-c^2\log(n)}\\
=\;&\frac{w\log^2 k}{w\log^2 n}+O\left(\frac{1}{\log n}\right)=\frac{\log^2 k}{\log^2 n}+O\left(\frac{1}{\log n}\right)
\end{aligned}
$$

Pour les convergences du type,

$$
\frac{1}{g^{1/2}(n)}\left(b_n-f(n)+\sum_{r=1}^{K}f(I_r^{(n)})\right)\xrightarrow{L_s}0, \tag{9.22}
$$

il est nécessaire de regarder de plus près l'effet des fonctions périodiques. Posons

$$
Z=\frac{1}{g^{1/2}(n)}\left(b_n-f(n)+\sum_{r=1}^{K}f(I_r^{(n)})\right). \tag{9.23}
$$

Les coefficients Z_1 et Z_2 sont donnés respectivement par

$$
\begin{aligned}
Z_1 = & \left(\frac{w_n \log(n) + \sqrt{d_n}}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right. \\
& \left. - \frac{c_n}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right) \mu_{0,k}\, k \\
& - \frac{c_n\left(\mu_{1,k}\, k \log(k) + \mu_{2,k}\, k\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& + \left(\frac{w_n \log(n) + \sqrt{d_n}}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right. \\
& \left. - \frac{c_n}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right) \mu_{0,n-k}\,(n-k) \\
& - \frac{c_n\left(\mu_{1,n-k}\,(n-k) \log(n-k) + \mu_{2,n-k}\,(n-k)\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& - \frac{\left(w_n \log(n) + \sqrt{d_n}\right)\left(\mu_{0,n}\, n - 1\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& + \frac{c_n\left(\mu_{1,n}\, n \log(n) + \mu_{2,n}\, n\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)},
\end{aligned}
$$

et

$$
\begin{aligned}
Z_2 = & \left(- \frac{c_n}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right. \\
& \left. + \frac{v_n + \sqrt{d_n}}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right) \mu_{0,k}\, k \\
& + \frac{\left(v_n + \sqrt{d_n}\right)\left(\mu_{1,k}\, k \log(k) + \mu_{2,k}\, k\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& + \left(- \frac{c_n}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right. \\
& \left. + \frac{v_n + \sqrt{d_n}}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \right) \mu_{0,n-k}\,(n-k) \\
& + \frac{\left(v_n + \sqrt{d_n}\right)\left(\mu_{1,n-k}\,(n-k) \log(n-k) + \mu_{2,n-k}\,(n-k)\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& + \frac{c_n\left(\mu_{0,n}\, n - 1\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)} \\
& - \frac{\left(v_n + \sqrt{d_n}\right)\left(\mu_{1,n}\, n \log(n) + \mu_{2,n}\, n\right)}{\log(n)\left(v_n w_n \log(n) + v_n\sqrt{d_n} + \sqrt{d_n} w_n \log(n) + d_n - c_n^2\right)}.
\end{aligned}
$$

En inspectant ces expressions, on voit que, pour montrer que Z tend vers zéro, il suffit de montrer, toujours dans L_s,

$$
\frac{1}{\sqrt{n}\log n}\left(\mu_{0,k} k + \mu_{0,n-k}(n-k) - \mu_{0,n} n\right) \longrightarrow 0, \tag{9.24}
$$

$$\frac{1}{\sqrt{n}\log^2 n}\left(\mu_{1,n}n\log n-\mu_{1,k}k\log k-\mu_{1,n-k}(n-k)\log(n-k)-\mu_{0,k}k-\mu_{0,n-k}(n-k)\right)\longrightarrow 0, \tag{9.25}$$

et

$$\frac{1}{\sqrt{n}\log n}\left(\mu_{2,n}n-\mu_{2,k}k-\mu_{2,n-k}(n-k)\right)\longrightarrow 0. \tag{9.26}$$

On peut montrer (9.24) et (9.26) comme dans la preuve du théorème 9.8. Pour montrer (9.25), écrivons :

$$\begin{aligned}
&\mu_{1,n}n\log n-\mu_{1,k}k\log k-\mu_{1,n-k}(n-k)\log(n-k)-\mu_{0,k}k-\mu_{0,n-k}(n-k)\\
=\;&\mu_{1,n}n\log n-\mu_{1,k}k\log k-\mu_{1,n-k}(n-k)\log(n-k)-\mu_{0,n}n && (9.27)\\
&+\mu_{0,n}n-\mu_{0,k}k-\mu_{0,n-k}(n-k). && (9.28)
\end{aligned}$$

On a

$$\frac{1}{\sqrt{n}\log^2 n}\left(\mu_{0,n}n-\mu_{0,k}k-\mu_{0,n-k}(n-k)\right)\longrightarrow 0$$

en procédant comme dans la preuve du théorème 9.8.

Il reste donc à montrer que le terme (9.27) tend vers zéro dans L_3. On va distinguer les cas $\dfrac{\log p}{\log q}$ rationnel et irrationnel.

Cas $\log p/\log q$ *rationnel.* —

Si $\dfrac{\log p}{\log q}\in\mathbb{Q}$, alors on peut écrire

$$\mu_{0n}n=\sum_{j=1}^{M}n^{-r_j}\alpha_j\varpi_{1j}(\log_\gamma n)$$

et

$$\mu_{1n}n=\sum_{j=1}^{M}n^{-r_j}\alpha_j h(r_j)\varpi_{1j}(\log_\gamma n).$$

On en déduit

$$\mu_{1,n} n \log n - \mu_{1,k} k \log k - \mu_{1,n-k}(n-k)\log(n-k) - \mu_{0,n} n \tag{9.29}$$

$$= \sum_{j=1}^{M} n^{-r_j} \log n \cdot \alpha_j h(r_j) \varpi_{1j}(\log_\gamma n) - \sum_{j=1}^{M} k^{-r_j} \log k \cdot \alpha_j h(r_j) \varpi_{1j}(\log_\gamma k) \tag{9.30}$$

$$- \sum_{j=1}^{M} (n-k)^{-r_j} \log(n-k) \cdot \alpha_j h(r_j) \varpi_{1j}(\log_\gamma (n-k)) \tag{9.31}$$

$$- \sum_{j=1}^{M} \alpha_j n^{-r_j} \cdot \alpha_j \varpi_{1j}(\log_\gamma n) \tag{9.32}$$

$$= \sum_{j=1}^{M} \alpha_j \varpi_{1j}(\log_\gamma n) \left(n^{-r_j} \log n - k^{-r_j} \log k - (n-k)^{-r_j} \log(n-k) - h(r_j) n^{-r_j} \right) \tag{9.33}$$

$$+ \sum_{j=1}^{M} \alpha_j k^{-r_j} \log k \left(\varpi_{1j}(\log_\gamma k) - \varpi_{1j}(\log_\gamma n) \right) \tag{9.34}$$

$$+ \sum_{j=1}^{M} \alpha_j (n-k)^{-r_j} \log(n-k) \left(\varpi_{1j}(\log_\gamma (n-k)) - \varpi_{1j}(\log_\gamma n) \right) \tag{9.35}$$

On a

$$\frac{1}{\sqrt{n} \log^2 n} \left(n \log n - k \log k - (n-k)\log(n-k) - hn \right) \longrightarrow 0$$

en utilisant le lemme 9.8 en-dessous. On montre de même

$$\frac{1}{\sqrt{n} \log^2 n} \left(n^{-r_j} \log n - k^{-r_j} \log k - (n-k)^{-r_j} \log(n-k) - h(r_j) n^{-r_j} \right) \longrightarrow 0$$

pour tout j en utilisant le lemme plus général 9.9.

Il reste à montrer des convergences du type

$$\frac{1}{\sqrt{n} \log^2 n} \left(k^{-r_j} \log k \left(\varpi_{1j}(\log_\gamma k) - \varpi_{1j}(\log_\gamma n) \right) \right) \longrightarrow 0$$

et

$$\frac{1}{\sqrt{n} \log^2 n} \left((n-k)^{-r_j} \log(n-k) \left(\varpi_{1j}(\log_\gamma (n-k)) - \varpi_{1j}(\log_\gamma n) \right) \right) \longrightarrow 0.$$

Or on a vu qu'il existe $\Phi_{n1}, \Psi_{n1}, \Phi_{n2}, \Psi_{n2}$ bornés tels que

$$\varpi_{1j}(\log_\gamma k) - \varpi_{1j}(\log_\gamma n) = \Phi_{n1} \Psi_{n1} \frac{k - np}{n}$$

et

$$\varpi_{1j}(\log_\gamma (n-k)) - \varpi_{1j}(\log_\gamma n) = \Phi_{n2} \Psi_{n2} \frac{n - k - nq}{n}.$$

En introduisant à nouveau les ensembles

$$E_n = \left\{ \left| I_1^{(n)} - np \right| < n^{3/4} \right\} = \left\{ \left| I_2^{(n)} - nq \right| < n^{3/4} \right\}$$

et

$$F_n = \left\{ \left| I_1^{(n)} - np \right| \geqslant n^{3/4} \right\} = \left\{ \left| I_2^{(n)} - nq \right| \geqslant n^{3/4} \right\},$$

en utilisant les bornes de Chernoff et en faisant usage des mêmes manipulations que dans la preuve du lemme 9.6 ci-dessous, on montre aisément les convergences

$$\frac{1}{\log n}\Phi_{n1}\Psi_{n1}\frac{k^{-r_j}}{n}\frac{\log k}{\log n}\frac{k-np}{\sqrt{n}} \longrightarrow 0$$

et

$$\frac{1}{\log n}\Phi_{n2}\Psi_{n2}\frac{(n-k)^{-r_j}}{n}\frac{\log k}{\log n}\frac{np-k}{\sqrt{n}} \longrightarrow 0$$

dans L_3.

Cas $\log p/\log q$ irrationnel. —

Si $\dfrac{\log p}{\log q} \notin \mathbb{Q}$, alors on peut écrire

$$\mu_{0n}n = \sum_{j\in\mathbb{Z}}^{M} n^{-z_j}\beta_j$$

et

$$\mu_{1n}n = \sum_{j\in\mathbb{Z}}^{M} n^{-z_j}\beta_j h(z_j),$$

où les β_j décroissent exponentiellement avec j.

On montre la convergence vers zéro comme dans la preuve du théorème 9.8. On se fixe un ε, on coupe la somme

$$\begin{aligned}
&\mu_{1,n}n\log n - \mu_{1,k}k\log k - \mu_{1,n-k}(n-k)\log(n-k) - \mu_{0,n}n \\
= &\sum_{j\in\mathbb{Z}}^{M} n^{-z_j}\log n\cdot\beta_j h(z_j) - \sum_{j\in\mathbb{Z}}^{M} k^{-z_j}\log k\cdot\beta_j h(z_j) \\
&-\sum_{j\in\mathbb{Z}}^{M}(n-k)^{-z_j}\log(n-k)\cdot\beta_j h(z_j) - \sum_{j\in\mathbb{Z}}^{M} n^{-z_j}\log n\cdot\beta_j
\end{aligned}$$

en deux, si bien qu'on se ramène à prouver les convergences

$$\frac{1}{\sqrt{n}\log^2 n}\left(n^{-z_j}\log n - k^{-z_j}\log k - (n-k)^{-z_j}\log(n-k) - h(z_j)n^{-z_j}\right) \longrightarrow 0$$

pour un nombre fini de racines z_j. □

9.3.3 Lemmes restants

Il reste maintenant à prouver les quelques résultats utilisés dans la preuve du théorème précédent, matérialisés par des lemmes, qui se démontrent essentiellement en ayant recours aux bornes de Chernoff. Montrer ces résultats demandent beaucoup de soin, et donc beaucoup de place, ce qui explique que par souci de lisibilité nous n'ayons pas intégré leur preuve dans la démonstration du théorème.

Lemme 9.6 *Soit $k \sim \mathcal{B}(n,p)$. Alors*

$$\sqrt{\frac{k}{n}}\frac{\log k}{\log n} \to p$$

et

$$\sqrt{\frac{k}{n}}\left(\frac{\log k}{\log n}\right)^2 \to p$$

dans tout L_s (s>2).

Démonstration. Il suffit de se rappeler que les fonctions $x \mapsto \sqrt{x}\log x$ et $x \mapsto \sqrt{x}\log^2 x$ sont bornées sur $[0,1]$, et d'écrire

$$\frac{\sqrt{k}\log k}{\sqrt{n}\log n} = \sqrt{\frac{k}{n}}\frac{\log n + \log\frac{k}{n}}{\log n} = \sqrt{\frac{k}{n}}\left(1+\frac{\log\frac{k}{n}}{\log n}\right) = \sqrt{\frac{k}{n}} + O\left(\frac{1}{\log n}\right)$$

et

$$\frac{\sqrt{k}\log^2 k}{\sqrt{n}\log^2 n} = \sqrt{\frac{k}{n}}\left(1+\frac{\log\frac{k}{n}}{\log n}\right)^2 = \sqrt{\frac{k}{n}} + O\left(\frac{1}{\log n}\right).$$

□

Lemme 9.7 *Soit $k \sim \mathcal{B}(n,p)$. Alors*

$$\sqrt{\frac{k}{n}}\frac{\log n}{\log k} \to p$$

dans tout L_s (s>2).

Démonstration. On écrit

$$\sqrt{\frac{k}{n}}\frac{\log n}{\log k} = \sqrt{\frac{k}{n}} + \frac{1}{\log k}\sqrt{\frac{k}{n}}\log\frac{k}{n}.$$

Or $\sqrt{\frac{k}{n}}\log\frac{k}{n}$ est borné, donc il suffit de montrer $\frac{1}{\log k} \to 0$ dans tout L_s. Si on pose

$$E_n = \left\{|k-np| < n^{3/4}\right\},$$

et $F_n = \complement E_n$, alors sur F_n, si on excepte le cas de probabilité asymptotiquement nulle où $k=1$, $|1/\log k|$ est majoré par $|1/\log 2|$, et pour tout $s>0$

$$\int_{F_n} |\log^{-s}| < \mu(F_n)/|\log 2| \longrightarrow 0,$$

puisqu'on a vu dans la preuve du théorème 9.8 que par les bornes de Chernoff

$$\mu(F_n) \leqslant 2e^{-\sqrt{n}/(2p)}.$$

D'autre part, pour n assez grand ($n > p^{-4}$),

$$\int_{E_n} |log^{-s}| < \frac{1}{np - n^{3/4}} \longrightarrow 0.$$

□

Lemme 9.8 *Soit $k \sim \mathcal{B}(n,p)$, et $h = p\log\frac{1}{p} + q\log\frac{1}{q}$. Alors*

$$\frac{1}{\sqrt{n\log^2 n}}\left(n\log n - k\log k - (n-k)\log(n-k) - hn\right) \xrightarrow{L_3} 0.$$

Démonstration. On commence par écrire

$$\begin{aligned} n\log n - k\log k - (n-k)\log(n-k) - hn &= n\left(p\log p + q\log q - \frac{k}{n}\log\frac{k}{n} - \frac{n-k}{n}\log\frac{n-k}{n}\right) \\ &= n\left[\left(p-\frac{k}{n}\right)\log p + \left(\log p - \log\frac{k}{n}\right)\frac{k}{n}\right. \\ &\quad \left. + \left(q - \frac{n-k}{n}\right)\log q + \left(\log q - \log\frac{n-k}{n}\right)\frac{n-k}{n}\right], \end{aligned}$$

puis, posant $I_1 = k$ et $I_2 = n-k$, on procède comme dans la preuve de (9.5) en introduisant la fonction $\eta(x) = \log\frac{x}{n}$, et c_r entre I_r et $n\pi_r$ si $|I_r - n\pi_r| < n^{3/4}$, tel que

$$\eta(I_r) = \eta(n\pi_r) + \eta'(c_r)(I_r - n\pi_r).$$

On pose $\Psi_{nr} = n\eta'(c_r)$, et il existe $M > 0$ tel que

$$\left|\Psi_{nr} - \frac{1}{\pi_r}\right| < Mn^{-1/4}.$$

On obtient

$$\begin{aligned} I_1\log\frac{I_1}{n} + I_2\log\frac{I_2}{n} + hn &= I_1\left(\log p + \Psi_{n1}\frac{I_1 - np}{n}\right) + I_2\left(\log p + \Psi_{n2}\frac{I_2 - np}{n}\right) + hn \\ &= (I_1 - np)\log p + (I_2 - nq)\log q + \Psi_{n1}I_1\frac{I_1 - np}{n} + \Psi_{n2}I_2\frac{I_2 - nq}{n} \\ &= (I_1 - np)\log\frac{p}{q} + \Psi_{n1}I_1\frac{I_1 - np}{n} + \Psi_{n2}I_2\frac{I_2 - nq}{n}. \end{aligned}$$

On a $\left|\frac{I_1 - np}{\sqrt{npq}}\right|^3 \sim |\mathcal{N}(0,1)|^3$, donc si $E_n = \left\{|I_1 - np| < n^{3/4}\right\}$, on a

$$\frac{1}{\log^2 n}\int_{E_n}\frac{I_1 - np}{\sqrt{n}} \xrightarrow{L_3} 0.$$

Il est clair que, si F_n est le complémentaire de E_n,

$$\frac{1}{\log^2 n}\int_{F_n}\frac{I_1 - np}{\sqrt{n}} \xrightarrow{L_3} 0.$$

Comme $\log\frac{p}{q}$, $\Psi_{n1}\frac{I_1}{n}$ et $\Psi_{n2}\frac{I_2}{n}$ sont bornés, cela assure la convergence vers 0 de chacun des termes et donc de la somme. □

Lemme 9.9 *Soient $k \sim \mathcal{B}(n,p)$, $h : t \mapsto -p^{-t}\log p - q^{-t}\log q$ et r une racine de $1 - p^{-s} - q^{-s}$. Alors*

$$\frac{1}{\sqrt{n\log^2 n}}\left(n^{-r}\log n - k^{-r}\log k - (n-k)\log(n-k) - h(r)n^{-r}\right) \xrightarrow{L_3} 0.$$

Démonstration. On commence par écrire

$$\begin{aligned}
& n^{-r}\log n - k^{-r}\log k - (n-k)^{-r}\log(n-k) - h(r)n^{-r} \\
= \; & n^{-r}\log n - k^{-r}\log k - (n-k)^{-r}\log(n-k) + \left(p^{-t}\log p + q^{-t}\log q\right)n^{-r} \\
= \; & n^{-r}\left(p^{-r}\log p + q^{-r}\log q - \left(\frac{k}{n}\right)^{-r}\log\frac{k}{n} - \left(\frac{n-k}{n}\right)^{-r}\log\frac{n-k}{n}\right. \\
& \left. + \left(1 - \left(\frac{k}{n}\right)^{-r} - \left(\frac{n-k}{n}\right)^{-r}\right)\log n\right) \\
= \; & n\left(p\log p + q\log q - \frac{k}{n}\log\frac{k}{n} - \frac{n-k}{n}\log\frac{n-k}{n}\right) \\
= \; & n\left[\left(p - \frac{k}{n}\right)\log p + \left(\log p - \log\frac{k}{n}\right)\frac{k}{n}\right. \\
& \left. + \left(q - \frac{n-k}{n}\right)\log q + \left(\log q - \log\frac{n-k}{n}\right)\frac{n-k}{n}\right].
\end{aligned}$$

On montre d'une part

$$\frac{1}{\sqrt{n\log^2 n}}n^{-r}\left(p^{-r}\log p + q^{-r}\log q - \left(\frac{k}{n}\right)^{-r}\log\frac{k}{n} - \left(\frac{n-k}{n}\right)^{-r}\log\frac{n-k}{n}\right) \xrightarrow{L_3} 0$$

en effectuant des calculs similaires à ceux de la preuve du lemme précédent.

D'autre part, en utilisant l'égalité $1 - p^{-r} - q^{-r} = 0$ et les bornes de Chernoff, on montre

$$\frac{1}{\sqrt{n\log^2 n}}n^{-r}\left(1 - \left(\frac{k}{n}\right)^{-r} - \left(\frac{n-k}{n}\right)^{-r}\right)\log n\right) \xrightarrow{L_3} 0.$$

□

9.4 Conclusion

Nous avons établi la limite gaussienne pour le nombre d'occurrences d'un trie fixé $\mathfrak{m}$ dans l'ensemble des tries avec une source de Bernoulli. Il reste alors à étudier l'effet de la compactification sur les tries, comme dans [88] ou [73].

Nous avons également montré une limite gaussienne pour la longueur de cheminement des tries, avec une moyenne en $n\log n$ et une variance en $n\log^2 n$ dans tous les cas, phénomène différent de la longueur de cheminement externe, d'espérance linéaire, et de variance linéaire si $p = q = \frac{1}{2}$, en $n\log n$ si $p \neq q$. Nous avons dû pour cela utiliser des méthodes de point fixe : l'analyse complexe bivariée semble insuffisante *a priori*.

Notons que nous avons dû établir quelques résultats intermédiaires pour pouvoir nous placer dans de bonnes hypothèses pour la dépoissonisation.

Une question mathématique consisterait à étudier la longueur de cheminement totale des tries, voire la somme des carrés ou d'une puissance quelconque des distances des nœuds à la racine. Pour ce dernier problème, des méthodes de point fixe ferait intervenir des matrices de plus grande dimension, avec vraisemblablement des termes en $n\log^d n$. Il faudrait donc trouver une manière de trouver l'asymptotique des coefficients des racines carrées de ces matrices de dimension supérieure à 2.

Conclusion

Dans cette thèse, nous avons étudié divers paramètres d'arbres, avec des méthodologies différentes suivant les problèmes, notamment suivant la nature des paramètres étudiés et des arbres considérés.

L'approche par fonctions génératrices suivie d'une analyse de singularité se montre très efficace pour calculer les premiers moments, et permet souvent d'obtenir la nature de la loi limite, notamment par théorème des quasi-puissances ou par pompage de moments.

La seule utilisation de l'analyse complexe se révèle parfois insuffisante pour conclure quant à la nature de la loi limite, auquel cas il faut parfois avoir recours à d'autres méthodes, commes les théorèmes de point fixe ou le mouvement brownien.

Des méthodes comme les martingales, qui paraissaient réservées aux arbres binaires de recherche, ont pu nous l'avons vu s'appliquer à des arbres de Catalan. Sans doute d'autres méthodologies pourront également élargir leur champ d'applications.

Les applications pratiques de l'étude de paramètres d'arbres n'ont pas été négligés puisque nous avons abordé entre autres la simplification d'expressions arithmétiques ou booléennes, la prédiction de la forme des structures secondaires d'ARN et la compactification des tries. Ces pistes de recherche mériteraient d'être explorées plus avant.

Annexes

Annexe A

Analyse de singularité

Cette annexe contient quelques rappels d'analyse de singularité. Rappelons que si A est une série $A(z) = \sum_n a_n z^n$, alors la notation $[z^n]A(z)$ désigne le coefficient a_n.

A.1 Analyse de singularité univariée

Dans [81], on trouve :

Théorème A.1 *Soit α un réel n'appartenant pas à $-\mathbb{N}$. Alors :*

$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+\sum_1^\infty \frac{e_k(\alpha)}{n^k}\right),$$

où $e_k(\alpha)$ est un polynôme en α de degré $2k$.

On a en particulier

$$[z^n] - \sqrt{1-z} \sim \frac{1}{2\sqrt{\pi n^3}}, \tag{A.1}$$

$$[z^n]\frac{1}{\sqrt{1-z}} \sim \frac{1}{\sqrt{\pi n}}. \tag{A.2}$$

Définition A.1 *Étant donné deux nombres ϕ, R avec $R > 1$ et $0 < \phi < \frac{\pi}{2}$, le domaine ouvert $\Delta(\phi, R)$ est défini par*

$$\Delta(\phi, R) = \left\{ z \;\middle|\; |z| < R,\ z \neq 1,\ |Arg(z-1)| > \phi \right\}.$$

Un domaine est un Δ-domaine si c'est un $\Delta(\phi, R)$ pour un certain $R > 1$ et un certain $\phi \in]0, \frac{\pi}{2}[$. Une fonction est Δ-analytique si elle est analytique dans un Δ-domaine.

Théorème A.2 (Transfert, première espèce) *Supposons que $f(z)$ est analytique et que, dans l'intersection d'un voisinage de 1 et d'un Δ-domaine, elle satisfait*

$$f(z) = O\left((1-z)^{-\alpha}\left(\log\frac{1}{1-z}\right)^\beta\right).$$

Alors

$$[z^n]f(z) = O\left(n^{\alpha-1}(\log n)^\beta\right).$$

Théorème A.3 (Transfert, deuxième espèce) *Supposons que $f(z)$ est analytique et que, dans l'intersection d'un voisinage de 1 et d'un Δ-domaine, elle satisfait*

$$f(z) = o\left((1-z)^{-\alpha}\left(\log\frac{1}{1-z}\right)^{\beta}\right).$$

Alors

$$[z^n]f(z) = o\left(n^{\alpha-1}(\log n)^{\beta}\right).$$

A.2 Limite discrète

Il peut arriver qu'une variable aléaoitre X_n à valeurs dans $\mathbb{N}$ converge vers une variable X sans normalisation. On parle alors de limite discrète. Plus formellement, on a la définition suivante [85] :

Définition A.2 *Une variable aléatoire X_n possède une limite discrète X de distribution $\{p_k\}_{k\geqslant 0}$ si*

$$\lim_{n\to\infty}\mathbf{Pr}[X_n = k] = p_k.$$

On parle alors de convergence en distribution.

A.3 Théorème des quasi-puissances

Dans [85], à propos du schéma algébraïco-logarithmique, on trouve le théorème suivant :

Théorème A.4 (Hwang) *Soit $F(z,u)$ une fonction bivariée analytique dans un domaine*

$$\mathcal{D}_0 = \{(z,u) \mid |z| < \rho, |u| < 1\},$$

dont le développement en série en (0,0) possède des coefficients positifs. Supposons qu'il existe $\varepsilon > 0$, $\vartheta < \frac{\pi}{2}$ et $r > \rho$ tels que dans le domaine

$$\mathcal{D} = \{(z,u) \mid |z| < r,\ Arg(z-\rho) \in [\vartheta, 2\pi - \vartheta],\ |u-1| < \varepsilon\},$$

la fonction admette la représentation

$$F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}\big(\log C(z,u)\big)^{k},$$

où $A(z,u)$, $B(z,u)$, $C(z,u)$ sont analytiques pour $(z,u) \in \mathcal{D}$, k est un entier positif, et $\alpha \notin \{0,-1,-2,\cdots\}$. Supposons également que l'équation $C(\zeta,1) = 0$ possède seulement une racine (simple) $\zeta = \rho$ dans $|z| \leqslant r$ et que $B(\rho,1) \neq 0$. Supposons finalement vérifiée la "condition de variabilité",

$$0 < \liminf \frac{\sigma_n^2}{n}.$$

Alors la variable avec fonction génératrice de probabilité $p_n(u) = \dfrac{[z^n]F(z,u)}{[z^n]F(z,1)}$ converge en distribution vers une variable gaussienne avec une vitesse de convergence en $O(n^{-1/2})$. La moyenne μ_n et la déviation standard σ_n sont asymptotiquement linéaires en n.

Ce théorème est très puissant et permet d'obtenir automatiquement la limite gaussienne de paramètres définis sur des objets combinatoires dès qu'on en connaît une série génératrice et que celle-ci a une expression sympathique. Ainsi par exemple, dans [170], j'ai pu rapidement trouver la série génératrice des oreilles de triangulations, et partant le caractère gaussien de la distribution limite, dont on ne connaissait que la moyenne asymptotiquement par des arguments combinatoires, qui faisait l'objet de l'article [107].

A.4 Remarques heuristiques

Lorsqu'on se trouve dans les hypothèses du théorème des quasi-puissances, on peut également faire les remarques suivantes.

Remarque A.1 *Les égalités*

$$\mu_n = \frac{[z^n]F_u(z,1)}{[z^n]F(z,1)} \tag{A.3}$$

et

$$\sigma_n^2 = \frac{[z^n]F_{u^2}(z,1)}{[z^n]F(z,1)} + \frac{[z^n]F_u(z,1)}{[z^n]F(z,1)} - \left(\frac{[z^n]F_u(z,1)}{[z^n]F(z,1)}\right)^2 \tag{A.4}$$

nous assurent, après analyse de singularité (théorème A.1), que μ_n et σ_n^2 possèdent un développement dans un échelle de $\frac{1}{n}$. Autrement dit, il n'y a pas de terme en $n^{3/2}$, $n^{1/2}$, $n^{-1/2}$, ni de manière générale en $n^{-p/2}$ avec $p \in \mathbb{N}$.

Remarque A.2 *Si f est une série génératrice de probabilité, et si $g_n(u) = f^n(u)$, alors il est clair que*

$$\mathbf{Var}(g_n) = g_n''(1) + g_n'(1) - g_n'(1)^2 = n\left(f''(1) + f'(1) - f'(1)^2\right) = n\,\mathbf{Var}(f),$$

si bien qu'il n'y a pas de terme en n^2. Le théorème des quasi-puissances, qui nous affirme que la loi limite est essentiellement donnée par $\left(\frac{1}{\rho(u)}\right)^n$, où $\rho(u)$ est la singularité en z et joue un rôle similaire à une série génératrice de probabilité, nous assure qu'il n'y a pas de terme en n^2 dans l'expansion en $1/n$ de

$$\sigma_n^2 = \frac{[z^n]F_{u^2}(z,1)}{[z^n]F(z,1)} + \frac{[z^n]F_u(z,1)}{[z^n]F(z,1)} - \left(\frac{[z^n]F_u(z,1)}{[z^n]F(z,1)}\right)^2.$$

Ces remarques sont pratiques lorsqu'on peut être amené à utiliser les formules (A.3) et (A.4) pour avoir des ordres de grandeur de la moyenne et de la variance à l'aide des seules séries génératrices, lorsque le calcul de $\rho(u)$ n'est *a priori* pas aisé.

A.5 Moments factoriels

Soit $F(z,u) = \sum_{n,k} f_{n,k} z^n u^k$ une fonction génératrice à deux variables, énumérant des objets combinatoires selon deux paramètres. Typiquement z énumère la taille de l'objet combinatoire.

L'ensemble des objets de taille n est énuméré par $F(z,1)$, si bien que la probabilité qu'un objet de taille n ait son second paramètre égal à est donné par

$$\frac{f_{nk}}{\sum_j f_{nj}} = \frac{[z^n u^k]F(z,u)}{[z^n]F(z,1)}.$$

Soit $m \in \mathbb{N}$. Le moment factoriel d'ordre m du paramètre énuméré par u, conditionné à ce que le paramètre énuméré par z, est donné par

$$\sum_k \frac{f_{nk}}{\sum_j f_{nj}} k(k-1)\cdots(k-m+1) = \frac{[z^n]F_{u^m}(z,1)}{[z^n]F(z,1)}.$$

Annexe B

Rappels sur le point fixe

Dans cette annexe, nous rappelons les principaux résultats sur les théorèmes de point fixe sur des distributions de probabilités.

On trouve les résultats cités ici dans [169]. Signalons également le livre [178] qui sert de référence sur les théorèmes de point fixe sur les distributions de probabilités, et contient tous les résultats importans accompagnés de leur preuve.

Pour $s = m + \alpha$ avec $0 < \alpha \leqslant 1$, $m \in \mathbb{N}_0$, on définit

$$\mathcal{F}_s := \left\{ f \in C^m(\mathbb{R}^d, \mathbb{R}) \;\middle|\; \left\| f^{(m)}(x) - f^{(m)}(y) \right\| \leqslant \|x - y\|^\alpha \right\}.$$

Zolotarev [215] a défini la métrique ζ_s sur des vecteurs de dimension d de la manière suivante :

$$\zeta_s = \sup_{f \in \mathcal{F}_s} |\mathbf{E}\,(f(X) - f(Y))|\,.$$

Théorème B.1 *La convergence selon ζ_s implique la convergence faible.*

Démonstration. On a convergence pour les fonctions cos et sin, d'où convergence pour la fonction caractéristique. □

Proposition B.1 *La loi normale est solution de l'équation*

$$X \stackrel{d}{=} \frac{1}{\sqrt{2}} X_1 + \frac{1}{\sqrt{2}} X_2,$$

où X_1 et X_2 sont des copies indépendantes de X.

On a une proposition similaire en dimension quelconque. Si on note ${}^tM$ la transposée d'une matrice M, alors on a la

Proposition B.2 *Soient A_1 et A_2 deux matrices de taille $d \times d$ telles que*

$$A_1 \cdot {}^tA_1 + A_2 \cdot {}^tA_2 = I_d,$$

où I_d est la matrice identité de dimension d. Alors la loi normale $\mathcal{N}(0, I_d)$ est solution de l'équation

$$X \stackrel{d}{=} A_1 \overline{X} + A_2 \overline{\overline{X}},$$

où $\overline{X}$ et $\overline{\overline{X}}$ sont des copies indépendantes de X.

Démonstration. Soit X de distribution $\mathcal{N}(0, I_d)$. Alors sa fonction caractéristique Φ_X vérifie

$$\Phi_X(u_1, \cdots, u_d) = \mathbf{E}\left[e^{i(\sum_{j=1}^d u_j X_j)}\right] = \mathbf{E}\left[e^{\frac{1}{2}i{}^tUU}\right],$$

donc pour toute matrice carrée A et tout vecteur U de dimension d,

$$\begin{aligned}
\Phi_{AX}(u_1, \cdots, u_d) &= \mathbf{E}\left[e^{i(\sum_{j=1}^d u_j (AX)_j)}\right] \\
&= \mathbf{E}\left[e^{i(\sum_{j=1}^d u_j (\sum_{k=1}^d a_{jk} X_k))}\right] \\
&= \mathbf{E}\left[e^{i(\sum_{k=1}^d \sum_{j=1}^d u_j (a_{jk} X_k))}\right] \\
&= \mathbf{E}\left[e^{i(\sum_{k=1}^d (\sum_{j=1}^d u_j a_{jk}) X_k)}\right] \\
&= \mathbf{E}\left[e^{i(\sum_{k=1}^d ({}^tUA)_k X_k)}\right] \\
&= \mathbf{E}\left[e^{i(\sum_{k=1}^d ({}^tAU)_k X_k)}\right] \\
&= \mathbf{E}\left[e^{\frac{1}{2}i{}^t({}^tAU){}^tAU}\right] \\
&= \mathbf{E}\left[e^{\frac{1}{2}i{}^tUA{}^tAU}\right],
\end{aligned}$$

et

$$\begin{aligned}
\Phi_{A_1\overline{X}+A_2\overline{\overline{X}}}(U) &= \mathbf{E}\left[e^{i(A_1\overline{X}U + A_2\overline{\overline{X}}U)}\right] \\
&= \mathbf{E}\left[e^{\frac{1}{2}i\ {}^tUA_1{}^tA_1U + \frac{1}{2}i\ {}^tUA_2{}^tA_2U}\right] \\
&= \mathbf{E}\left[e^{\frac{1}{2}i\ {}^tU(A_1{}^tA_1 + A_2{}^tA_2)U}\right] \\
&= \mathbf{E}\left[e^{\frac{1}{2}i\ {}^tUU}\right] \\
&= \Phi_X(U).
\end{aligned}$$

□

Les résultats suivants sont fournis sans preuve.

Proposition B.3 *La finitude de $\zeta_s(X, Y)$ implique que X et Y ont des moments identiques jusqu'à l'ordre m.*

Pour $2 < s \leqslant 3$, un vecteur $m \in \mathbb{R}^d$, soit

$$\mathcal{M}_s^d(m, C) := \left\{\mu \in \mathcal{M}^d : \|\mu\|_s < \infty, \mathbf{E}\mu = m, \mathbf{Cov}(\mu) = C\right\}.$$

Soit l'opérateur

$$T : \mathcal{M}_s^d(m, C) \longrightarrow \mathcal{M}_s^d(m, C), \quad \mu \longmapsto \mathcal{L}\left(A_1 Z^{(1)} + A_2 Z^{(2)}\right),$$

où A_1, A_2, $Z^{(1)}$ et $Z^{(2)}$ sont indépendants et $Z^{(r)} \sim \mu$ pour $r = 1, 2$.

Proposition B.4 *Si*

$$\xi := \mathbf{E}\left[\|A_1\|_{op}^s + \|A_2\|_{op}^s\right] < 1,$$

alors T possède un unique point fixe dans $\mathcal{M}_s^d(m, C)$.

Théorème B.2 (Théorème de transfert univarié) *Soit Y_n vérifiant, pour un certain entier n_0,*

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} Y^{(r)}_{I_r^{(n)}} + b_n, \qquad n \geqslant n_0, \tag{B.1}$$

avec $(Y_n^{(1)}), \cdots, (Y_n^{(K)})$, $(I_r^{(n)}, b_n)$ indépendants, vérifiant

$$Y_j^{(r)} \sim Y_j, \text{ pour tout } j \geqslant 0,$$

$$\mathbf{Pr}[I_r^{(n)} = n] \underset{n\to\infty}{\longrightarrow} 0 \text{ pour } r = 1, \cdots, K,$$

et il existe n_1 tel que

$$Var(Y_n) > 0 \text{ pour } n \geqslant n_1.$$

Supposons que pour les fonctions $f, g : \mathbb{N} \to \mathbb{R}_0^+$ avec $g(n) > 0$ pour n suffisamment grand nous ayons d'une part la condition de stabilité *dans L_s pour $s \in]0, 3]$, qui se traduit par les convergences*

$$\frac{1}{g^{1/2}(n)} \left(b_n - f(n) + \sum_{r=1}^{K} f(I_r^{(n)}) \right) \xrightarrow{L_s} 0 \tag{B.2}$$

et

$$\left(\frac{g(I_r^{(n)})}{g(n)} \right)^{1/2} \xrightarrow{L_s} A_r^*, \qquad r = 1, \cdots, K, \tag{B.3}$$

et d'autre part la condition de contraction

$$\mathbf{E} \sum_{r=1}^{K} (A_r^*)^s < 1.$$

Supposons

$$\begin{aligned} &\mathbf{E}Y_n = f(n) + o(g^{1/2}(n)), \ \mathbf{Var}(Y_n) = g(n) + o(g(n)), \quad si \ \ 2 < s \leqslant 3, \\ &\mathbf{E}Y_n = f(n) + o(g^{1/2}(n)), \quad si \ \ 1 < s \leqslant 2. \end{aligned}$$

Alors

$$\frac{Y_n - f(n)}{g^{1/2}(n)} \xrightarrow{\mathcal{L}} X,$$

où X est l'unique point fixe de

$$X \stackrel{d}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*$$

dans $\mathcal{M}_s^1(0,1)$, et $(A_1^*, \cdots, A_K^*, b^*)$, $X^{(1)}, \cdots, X^{(K)}$ sont indépendants avec $X^{(r)} \sim X$.

Corollaire B.1 *Supposons que Y_n vérifie (B.1), et qu'il existe de plus $s \in]2,3]$ tel que*

$$\frac{1}{g^{1/2}(n)}\left(b_n - f(n) + \sum_{r=1}^{K} f(I_r^{(n)})\right) \xrightarrow{L_s} 0 \tag{B.4}$$

et

$$\left(\frac{g(I_r^{(n)})}{g(n)}\right) \xrightarrow{L_s} A_r^*, \tag{B.5}$$

avec

$$\sum_{r=1}^{K}(A_r^*)^2 = 1,\ \mathbf{Pr}(\exists r,\ A_r^* = 1) < 1. \tag{B.6}$$

Alors

$$\frac{Y_n - f(n)}{g^{1/2}(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

où $\mathcal{N}(0,1)$ désigne la distribution normale standard.

Bibliographie

[1] Abramowitz (Milton) and Stegun (Irene A.). – *Handbook of Mathematical Functions.* – Dover, 1973. A reprint of the tenth National Bureau of Standards edition, 1964.

[2] Aho (Alfred V.), Sethi (Ravi), and Ullman (Jeffrey D.). – *Compilers. Principles, Techniques, and tools.* – Addison-Wesley, 1988.

[3] Albert (Luc), Casas (Rafael), and Fages (François). – Average-case analysis of unification algorithms. *Theoretical Computer Science*, vol. 113, n°1, 1993, pp. 3–34.

[4] Baccelli (François), Cohen (Guy), Olsder (Geert Jan), and Quadrat (Jean-Pierre). – *Synchronization and Linearity.* – John Wiley & Sons, Paris, 1992.

[5] Baeza-Yates (Ricardo), Casas (Rafael), Díaz (Josep), and Martínez (Conrado). – On the average size of the intersection of binary trees. *SIAM J. on Computing*, vol. 21, n°1, 1992, pp. 24–32.

[6] Banderier (Cyril), Flajolet (Philippe), Schaeffer (Gilles), and Soria (Michèle). – Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures & Algorithms*, vol. 19, n°3/4, 2001, pp. 194–246.

[7] Barbe (Philippe) and Ledoux (Michel). – *Probabilités.* – Belin, 1998, *De la licence à l'agrégation.* 262 pages.

[8] Bender (Edward Anton). – Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory*, vol. 15, 1973, pp. 91–111.

[9] Bender (Edward Anton). – Asymptotic methods in enumeration. *SIAM Review*, vol. 16, n°4, October 1974, pp. 485–515.

[10] Bender (Edward Anton), Richmond (Lawrence Bruce), and Wormald (Nicholas Charles). – Largest 4-connected components of 3-connected planar triangulations. *Random Structures and Algorithms*, vol. 7, n°4, 1995, pp. 273–285.

[11] Bergeron (François), Labelle (Gilbert), and Leroux (Pierre). – *Combinatorial Species and Tree-like Structures.* – Cambridge University Press, 1998, *Encyclopedia of Mathematics and its Applications*, vol. 67, XX, 457 s.p.

[12] Berstel (Jean) and Reutenauer (Christophe). – Recognizable formal power series on trees. *Theoretical Computer Science*, vol. 18, 1982, pp. 115–148.

[13] Bertoin (Jean) and Pitman (Jim). – Path transformations connecting brownian bridge, excursion and meander. *Bull. Sci. Math. (2)*, vol. 118, 1994, pp. 147–166.

[14] Biane (Philippe) and Yor (Marc). – Valeurs principales associées au temps locaux Browniens. *Bull. Sci. Math. (2)*, vol. 111, 1987, pp. 23–101.

[15] Billingsley (Patrick). – *Probability and Measure, third edition.* – John Wiley & Sons, 1995.

[16] Blondel (Vincent D.). – Structured numbers : Properties of a hierarchy of operations on binary trees. *Acta Informatica*, vol. 35, n°1, 1998, pp. 1–15.

[17] Bourdon (Jérémie). – Size and Path Length of Patricia Tries : Dynamical Sources Context. *Random Structures and Algorithms*, vol. 19, n°3-4, 2001, pp. 289–315.

[18] Bourdon (Jérémie). – *Méthodes d'analyse dynamique dans l'étude en moyenne d'algorithmes arithmétiques et de la théorie de l'information.* – PhD thesis, Université de Caen, December 2002.

[19] Bousquet-Mélou (Mireille). – Une bijection entre les polyominos convexes dirigés et les mots de dyck bilatères. *R.A.I.R.O. Informatique Théorique et Applications*, vol. 26, 1992, pp. 205–219.

[20] Bousquet-Mélou (Mireille). – Codage des polyominos convexes et équations pour l'énumération suivant l'aire. *Discrete Applied Mathematics*, vol. 48, 1994, pp. 21–43.

[21] Bousquet-Mélou (Mireille). – *Rapport scientifique.* – Habilitation à diriger des recherches, Université Bordeaux I, December 1996.

[22] Breiman (Leo). – *Probability.* – SIAM, Philadelphia, 1992, *Classics in applied mathematics; 7.*

[23] Brémaud (Pierre). – *Introduction aux Probabilités.* – Springer, 1988.

[24] Bruccoleri (Robert) and Heinrich (Gerhard). – An improved algorithm for nucleic acid secondary structure display. *Comput. Appl. Biosci.*, vol. 4, 1988, pp. 167–173.

[25] Canfield (E. Rodney). – Remarks on an asymptotic method in combinatorics. *Journal of Combinatorial Theory*, Series A, vol. 37, 1984, pp. 348–352.

[26] Casas (Rafael), Fernández Camacho (María-Inés), and Steyaert (Jean-Marc). – Algebraic Simplification in Computer Algebra : An Analysis of Bottom-up Algorithms. *Theoretical Computer Science*, vol. 74, n°74, 1990, pp. 273–298.

[27] Casas (Rafael) and Steyaert (Jean-Marc). – Bottom-up recursion in trees. In *Proceedings of CAAP'86. Lecture Notes in Computer Science*, vol. 214, pp. 172–182. – Springer-Verlag, 1986.

[28] Casas (Ricardo), Díaz (Josep), and Martínez (Conrado). – Average-case analysis on simple families of trees using a balanced probability model. *Theoretical Comp. Sci.*, vol. 117, 1993, pp. 99–112.

[29] Chauvin (Brigitte). – Martingales discrètes et applications à l'analyse d'algorithmes. In Chyzak (Frédéric) (editor), *Algorithms seminar, 2001-2002.* pp. 151–166. – Research Report n°5003, INRIA, 2002.

[30] Chauvin (Brigitte), Drmota (Michael), and Jabbour-Hattab (Jean). – The profile of binary search trees. *Annals of Probability*, vol. 11, 2001, pp. 1042–1062.

[31] Chauvin (Brigitte), Flajolet (Philippe), Gardy (Danièle), and Gittenberger (Bernhard). – And/or trees revisited. *Combinatorics, Probability, and Computing*, 2002.

[32] Chern (Hua-Huai) and Hwang (Hsien-Kuei). – Phase changes in random m-ary search trees and generalized quicksort. *Random Structures and Algorithms*, vol. 19, 2001, pp. 316–358.

[33] Choppy (Christine), Kaplan (Stéphane), and Soria (Michèle). – Complexity analysis of term rewriting systems. *Theoretical Computer Science*, vol. 67, 1989, pp. 261–282.

[34] Chossat (Maurice). – *Mathématiques de l'ingénieur.* – Dunod, Paris, 2001, *Aide-mémoire.*

[35] Cifarelli (Donato Michele). – Contributi intorno ad un test per l'omogeneità tra due campioni. *G. Econom. Ann. Econ. (N.S.)*, vol. 34, 1975, pp. 233–249.

[36] Clément (Julien). – *Arbres Digitaux et Sources Dynamiques.* – Doctorat, Université de Caen, 1998.

[37] Clément (Julien), Flajolet (Philippe), and Vallée (Brigitte). – The analysis of hybrid trie structures. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, 1998).* pp. 531–539. – SIAM Press, Philadelphia, PA, 1998.

[38] Clément (Julien), Flajolet (Philippe), and Vallée (Brigitte). – Dynamical sources in information theory : A general analysis of trie structures. *Algorithmica*, vol. 29, 2001, pp. 307–369.

[39] Darboux (Gaston). – Mémoire sur l'approximation des fonctions de très grands nombres, et sur une classe étendue de développements en série. *Journal de Mathématiques Pures et Appliquées*, February 1878, pp. 5–56,377–416.

[40] Delest (Maylis P.) and Viennot (Gérard). – Algebraic languages and polyominoes enumeration. *Theoretical Computer Science*, vol. 34, 1984, pp. 169–206.

[41] Devroye (Luc). – A note on the height of binary search trees. *Journal of the ACM*, vol. 33, 1986, pp. 489–498.

[42] Devroye (Luc) and Kamoun (Olivier). – Random minimax game trees. In Aldous (David) and Pemantle (Robin) (editors), *Random Discrete Structures.* pp. 55–80. – John Wiley, New York, 1996.

[43] Devroye (Luc) and Kruszewski (Paul). – A note on the Horton-Strahler number for random binary trees. *Information Processing Letters*, vol. 52, 1991, pp. 155–159.

[44] Devroye (Luc), Neininger (Ralph), and Fill (James). – Perfect simulation from the quicksort limit distribution. *Electronic Communications in Probability*, vol. 5, 2000, pp. 95–99.

[45] Drmota (Michael). – Asymptotic distributions and a multivariate Darboux method in enumeration problems. *Journal of Combinatorial Theory*, Series A, vol. 67, 1994, pp. 169–184.

[46] Drmota (Michael). – Systems of functional equations. *Random Structures and Algorithms*, vol. 10, 1997, pp. 103–124.

[47] Drmota (Michael). – Stochastic analysis of tree-like data structures. *Proc. R. Soc. Lond.*, vol. A 460, 2004, pp. 271–307.

[48] Drmota (Michael) and Marckert (Jean-François). – Reinforced weak convergence of stochastic processes. *Statistics Probab. Letters*, 2004. – To appear.

[49] Drmota (Michael) and Soria (Michèle). – Marking in combinatorial constructions : Generating functions and limiting distributions. *Theoretical Computer Science*, vol. 144, n°1–2, 1995, pp. 67–99.

[50] Duchi (Enrica), Frosini (Andrea), Pinzani (Renzo), and Rinaldi (Simone). – A note on rational succession rules. *Journal of Integer Sequences*, vol. 6, 2003. – Article 03.1.7.

[51] Duchon (Philippe). – *Q-grammaires : un outil pour l'énumération.* – Doctorat, Université de Bordeaux I, 1998.

[52] Duchon (Philippe). – Q-grammars and wall polyominoes. *Annals of Combinatorics*, vol. 3, 1999, pp. 311–321.

[53] Duflo (Marie). – *Méthodes récursives aléatoires.* – Masson, Paris, 1990.

[54] Duflo (Marie) and Dacuhna-Castelle (Didier). – *Probabilités et Statistiques, tomes 1 et 2.* – Masson, Paris, 1984.

[55] Dulucq (Serge) and Tichit (Laurent). – À propos de la comparaison de structures secondaires d'ARN. In *Actes du Colloque JOBIM 2001 Biologie, Informatique Mathématiques*, pp. 125–134. – Toulouse, 2001. June 2001.

[56] Dulucq (Serge) and Tichit (Laurent). – RNA secondary structure comparison : exact analysis of the Zhang-Shasha tree edit algorithm. *Theoretical Computer Science*, 2003. – To appear.

[57] Durand (Marianne). – Asymptotic analysis of an optimized quicksort. *Inf. Process. Lett.*, vol. 85, n°2, 2003, pp. 73–77.

[58] Durand (Marianne) and Taylor (Stephen). – Emerging behavior as binary search trees are symmetrically updated. *Theoretical Computer Science*, vol. 297, n°1-3, 2003, pp. 425–445.

[59] Dutour (Isabelle). – *Grammaires d'objets : énumération, bijections et génération aléatoire.* – Doctorat, Université de Bordeaux I, 1996.

[60] Ethier (Stewart N.) and Kurtz (Thomas G.). – *Markov Processes : Characterization and Convergence.* – Wiley, New York, 1986.

[61] Fédou (Jean-Marc). – *Grammaires et q-énumération de polyominos.* – Doctorat, Université Bordeaux I, 1989.

[62] Feller (William). – *An Introduction to Probability Theory and Its Applications.* – John Wiley, 1968, third edition, vol. 1.

[63] Feller (William). – *An Introduction to Probability Theory and Its Applications.* – John Wiley, 1971, vol. 2.

[64] Fernández-Camacho (María-Inés). – *Analisis Medio de Algoritmos de reducción sobre árboles.* – Tesis, Universidad Complutense de Madrid, 1988.

[65] Fill (James), Flajolet (Philippe), and Kapur (Nevin). – Singularity analysis, hadamard products, and tree recurrence. *Journal of Computational and Applied Mathematics*, 2003. – Submitted.

[66] Fill (James) and Janson (Svante). – A characterization of the set of fixed points of the quicksort transformation. *Electronic Communications in Probability*, vol. 5, 2000, pp. 77–84.

[67] Fill (James) and Janson (Svante). – Smoothness and Decay Properties of the Limiting Quicksort Density Function. In *Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities. Trends in Mathematics*, pp. 53–64. – Birkhäuser, Versailles, 2000.

[68] Fill (James) and Janson (Svante). – Approximating the limiting quicksort distribution. *Special issue on Analysis of Algorithms, Random Structures & Algorithms*, vol. 19, 2001, pp. 376–406.

[69] Fill (James) and Janson (Svante). – Quicksort asymptotics. *Special issue on Analysis of Algorithms, Journal of Algorithms*, vol. 44, 2002, pp. 4–28.

[70] Flajolet (Philippe). – Singularity analysis and asymptotics of Bernoulli sums. *Theoretical Computer Science*, vol. 215, n°1-2, 1999, pp. 371–381.

[71] Flajolet (Philippe), Gao (Zhicheng), Odlyzko (Andrew), and Richmond (Bruce). – The distribution of heights of binary trees and other simple trees. *SIAM Journal on Discrete Mathematics*, vol. 2, 1993, pp. 145–156.

[72] Flajolet (Philippe), Gourdon (Xavier), and Dumas (Philippe). – Mellin transforms and asymptotics : harmonic sums. *Theoretical Computer Science*, Special Volume on Mathematical Analysis of Algorithms, vol. 144, nl-2, June 1995, pp. 3–58.

[73] Flajolet (Philippe), Gourdon (Xavier), and Martínez (Conrado). – Patterns in Random Binary Search Trees. *Random Structures and Algorithms*, vol. 11, n°3, October 1997, pp. 223–244.

[74] Flajolet (Philippe) and Odlyzko (Andrew M.). – The average height of binary trees and other simple trees. *Journal of Computer and System Sciences*, vol. 25, 1982, pp. 171–213.

[75] Flajolet (Philippe) and Odlyzko (Andrew M.). – Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, vol. 3, n°2, 1990, pp. 216–240.

[76] Flajolet (Philippe) and Prodinger (Helmut). – Register allocation for unary binary trees. *SIAM Journal on Computing*, vol. 15, n°3, August 1986, pp. 629–640.

[77] Flajolet (Philippe), Raoult (Jean-Claude), and Vuillemin (Jean). – The number of registers required for evaluating arithmetic expressions. *Theoretical Computer Science*, vol. 9, 1979, pp. 99–125.

[78] Flajolet (Philippe), Régnier (Mireille), and Sedgewick (Robert). – Some Uses of the Mellin Integral Transform in the Analysis of Algorithms. In Apostolico (A.) and Galil (Z.) (editors), *Combinatorial Algorithms on Words. NATO Advance Science Institute Series.* Series F : Computer and Systems Sciences, vol. 12, pp. 241–254. – Springer-Verlag, 1985.

[79] Flajolet (Philippe), Régnier (Mireille), and Sotteau (Dominique). – *Algebraic Methods for Trie Statistics.* – Research Report n°298, Institut National de Recherche en Informatique et en Automatique, 1984. 32 pages.

[80] Flajolet (Philippe), Régnier (Mireille), and Sotteau (Dominique). – Algebraic methods for trie statistics. *Annals of Discrete Mathematics*, vol. 25, 1985, pp. 145–188.

[81] Flajolet (Philippe) and Sedgewick (Robert). – *The Average Case Analysis of Algorithms : Complex Asymptotics and Generating Functions.* – Research Report n°2026, Institut National de Recherche en Informatique et en Automatique, 1993. 100 pages.

[82] Flajolet (Philippe) and Sedgewick (Robert). – *The Average Case Analysis of Algorithms : Counting and Generating Functions.* – Research Report n°1888, Institut National de Recherche en Informatique et en Automatique, 1993. 116 pages.

[83] Flajolet (Philippe) and Sedgewick (Robert). – Mellin transforms and asymptotics : finite differences and Rice's integrals. *Theoretical Computer Science*, vol. 144, n°1–2, June 1995, pp. 101–124.

[84] Flajolet (Philippe) and Sedgewick (Robert). – *The Average Case Analysis of Algorithms : Mellin Transform Asymptotics.* – Research Report n°2956, Institut National de Recherche en Informatique et en Automatique, 1996. 93 pages.

[85] Flajolet (Philippe) and Sedgewick (Robert). – *The Average Case Analysis of Algorithms : Multivariate Asymptotics and Limit Distributions.* – Research Report n°3162, Institut National de Recherche en Informatique et en Automatique, 1997. 123 pages.

[86] Flajolet (Philippe) and Sedgewick (Robert). – *Functional Equations, Rational and Algebraic Functions.* – Research Report n°4103, Institut National de Recherche en Informatique et en Automatique, 2001. 98 pages.

[87] Flajolet (Philippe) and Sedgewick (Robert). – *Analytic Combinatorics : Basic Complex Asymptotics.* – 2003. `http ://algo.inria.fr/flajolet/Publications/FlSe03.ps`. 133 pages.

[88] Flajolet (Philippe), Sipala (Paolo), and Steyaert (Jean-Marc). – Analytic variations on the common subexpression problem. In Paterson (M. S.) (editor), *Automata, Languages, and Programming, Lecture Notes in Computer Science*, vol. 443, pp. 220–234. – 1990. Proceedings of the 17th ICALP Conference, Warwick, July 1990.

[89] Flajolet (Philippe) and Soria (Michèle). – Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory,* Series A, vol. 53, 1990, pp. 165–182.

[90] Flajolet (Philippe) and Steyaert (Jean-Marc). – A complexity calculus for recursive tree algorithms. *Mathematical Systems Theory,* vol. 19, 1987, pp. 301–331.

[91] Foata (Dominique) and Fuchs (Aimé). – *Calcul des probabilités.* – Dunod, Paris, 1998.

[92] Fontana (W.), Konings (D.), Stadler (P.), and Schuster (P.). – Statistics of rna secondary structures. *Biopolymers*, vol. 33, 1993, pp. 1389–1404.

[93] Françon (Jean). – Sur le nombre de registres nécessaires à l'évaluation d'une expression arithmétique. *RAIRO Theoret Info Appl*, vol. 18, n°2, 1984, pp. 355–364.

[94] Gao (Zhicheng) and Wormald (Nicholas). – The size of the largest components in random planar graphs. *SIAM Journal of Discrete Mathematics*, vol. 12, n°2, 1999, pp. 217–228.

[95] Gardy (Danièle). – Méthode de col et lois limites en analyse combinatoire. *Theoretical Computer Science*, vol. 92, n°2, 1992, pp. 261–280.

[96] Gardy (Danièle). – Some results on the asymptotic behaviour of coefficients of large powers of functions. *Discrete Mathematics*, vol. 139, n°1–3, 1995, pp. 189–217.

[97] Gaspin (Christine) and Westhof (Éric). – The determination of the secondary structures of rna as a constraint satisfaction problem. In SCHULZE-KREMER S. (IOS Press) (editor), *Advances in Molecular Bioinformatics.* – 1994.

[98] Goulden (Ian P.) and Jackson (David M.). – *Combinatorial Enumeration.* – John Wiley, New York, 1983.

[99] Graham (Ronald L.), Knuth (Donald Ervin), and Patashnik (Oren). – *Concrete Mathematics.* – Addison Wesley, 1989.

[100] Grimmett (Geoffrey) and Stirzaker (David). – *Probability and Random Processes.* – Oxford University Press, 2001, third edition.

[101] Groeneboom (Piet). – Brownian motion with a parabolic drift and airy functions. *Probability Theory and Related Fields*, vol. 81, 1989, pp. 79–109.

[102] Gutjahr (Walter J.) and Pflug (Georg Ch.). – The asymptotic contour process of a binary tree is a brownian excursion. *Stochastic Processes and their Applications*, vol. 41, 1992, pp. 69–89.

[103] Hagerup (Torben) and Rüb (Christine). – A Guided Tour of Chernoff Bounds. *Information Processing Letters*, vol. 33, 1989, pp. 305–308.

[104] Hall (Peter G.) and Heyde (Christopher C.). – *Martingale limit theory and its application.* – Academic Press, New York, 1980.

[105] Hennequin (Pascal). – *Analyse en moyenne d'algorithmes, tri rapide et arbres de recherche.* – Doctorat, École Polytechnique, March 1994. 162 pages.

[106] Howell (JoAnn), Smith (Temple F.), and Waterman (Michael S.). – Computation of generating functions for biological molecules. *SIAM J. Appl. Math.*, vol. 39, 1980, pp. 119–133.

[107] Hurtado (Ferran) and Noy (Marc). – Ears of triangulations and Catalan numbers. *Discrete Mathematics*, vol. 149, 1996, pp. 319–324.

[108] Hwang (Hsien-Kuei). – *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* – Doctorat, École Polytechnique, December 1994.

[109] Hwang (Hsien-Kuei). – Phase change of limit laws in the quicksort recurrences under varying toll functions. *SIAM Journal on Computing*, vol. 31, 2002.

[110] Imhof (Jean-Pierre). – Density factorizations for brownian motion, meander and the three-dimensional bessel process, and applications. *Journal of Applied Probability*, vol. 21, 1984, pp. 500–510.

[111] Ito (Kiyosi) and McKean (Henry P. Jr.). – *Diffusion processes and their sample paths.* – Academic Press, New York, 1965.

[112] Jacod (Jean). – *Calcul stochastique et problèmes de martingales.* – Springer-Verlag, Berlin, Heidelberg, New York, 1979.

[113] Jacquet (Philippe) and Régnier (Mireille). – Trie partitioning process : Limiting distributions. In Franchi-Zanetacchi (P.) (editor), *CAAP'86, Lecture Notes in Computer Science*, vol. 214, pp. 196–210. – 1986. Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice France, March 1986.

[114] Jacquet (Philippe) and Régnier (Mireille). – *Normal limiting distributions for the size and the external path length of tries.* – Research Report n°827, Institut National de Recherche en Informatique et en Automatique, 1998. 18 pages.

[115] Janson (Svante). – Ideals in a forest, one-way infinite binary trees and the contraction method. In *Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities. Trends in Mathematics*, pp. 393–414. – Birkhäuser, 2002. Versailles.

[116] Jeanblanc (Monique), Pitman (Jim), and Yor (Marc). – The Feynman-Kac formula and decomposition of Brownian paths. *Computational and Applied Mathematics*, vol. 16, 1997, pp. 27–52.

[117] Juan (Vernica) and Wilson (Charles). – RNA secondary structure prediction based on free energy and phylogenetic analysis. *Journal of Molecular Biology*, vol. 289, 1999, pp. 935–947.

[118] Kac (Marc) and Erdös (Paul). – On certain limits theorems of the theory of probability. *Bulletin of the American Mathematical Society*, vol. 52, 1944, pp. 292–302.

[119] Kac (Mark). – On some connections between Probability Theory and Differential and Integral equations. In Neyman (J.) (editor), *Proc. 2nd Berkeley Symp. Math. Stat. Prob.*, pp. 189–215. – 1949. University of California.

[120] Kac (Mark). – On the distribution of certain Wiener functionals. *Transactions of the American Mathematical Society*, vol. 65, 1949, pp. 1–13.

[121] Kac (Mark). – *Probability and related topics in physical sciences.* – Interscience Publishers, London, New York, 1959. Summer seminar on applied mathematics (University of Colorado). (1; 1957; Boulder, Colorado).

[122] Karatzas (Ioannis) and Shreve (Steven E.). – *Brownian Motion and Stochastic Calculus, 2nd ed.* – Springer-Verlag, New York, 1997.

[123] Karlin (Samuel) and Taylor (Howard M.). – *A First Course in Stochastic Processes. Second edition.* – Academic Press, New York, 1997.

[124] Kemp (Rainer). – The average number of registers needed to evaluate a binary tree optimally. *Acta Informatica*, vol. 11, 1979, pp. 363–372.

[125] Kemp (Rainer). – *Fundamentals of the Average Case Analysis of Particular Algorithms.* – John Wiley & Sons, Inc., New York, 1985.

[126] Kingman (John F. C.). – Subadditive ergodic theory. *Annals of Probability*, vol. 1, 1973, pp. 883–909.

[127] Knessl (Charles) and Szpankowski (Wojciech). – Quicksort algorithm again revisited. *Discrete Mathematics and Theoretical Computer Science*, vol. 3, 1999, pp. 43–64.

[128] Knight (Frank B.). – On the excursion process of brownian motion. *Transactions of the American Mathematical Society*, vol. 258, 1980, pp. 77–86.

[129] Knuth (Donald Ervin). – *The Art of Computer Programming.* – Addison-Wesley, 1968, vol. 1 : Fundamental Algorithms. Second edition, 1973.

[130] Knuth (Donald Ervin). – Semantic of context-free languages. *Mathematical Systems theory*, vol. 2, 1968, pp. 127–145. – Correction : Mathematical Systems theory 5 :95-96, 1971.

[131] Knuth (Donald Ervin). – Mathematical analysis of algorithms. – IFIP Congress, Ljubljana, Yugoslavia, 1971.

[132] Knuth (Donald Ervin). – *The Art of Computer Programming.* – Addison-Wesley, 1973, vol. 3 : Sorting and Searching.

[133] Kozen (Dexter C.). – *Automata and computability.* – Springer, New York, Berlin, Paris, 1997, XIII–400p.

[134] Kreweras (Germain). – Traitement simultané du "Problème de Young" et du "Problème de Simon Newcomb". *Cahiers du Bureau Universitaire de Recherche Opérationnelle*, vol. 1, 1967, pp. 333–350.

[135] Lefebvre (Fabrice). – *Grammaires S-attribuées multi-bandes et applications à l'analyse automatique de séquences biologiques.* – PhD thesis, Université de Paris-Sud, Orsay, July 1990.

[136] Lefebvre (Fabrice). – An optimized parsing algorithm well-suited to RNA folding. In press (AAAI) (editor), *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*, pp. 222–230. – 1995.

[137] Lefebvre (Fabrice). – A grammar-based unification of several alignment and folding algorithms. In press (AAAI) (editor), *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology*, pp. 143–154. – 1996.

[138] Levinger (Bernard W.). – The square root of a 2×2 matrix. *Math. Mag.*, vol. 53, n °4, 1980, pp. 222–224.

[139] Lévy (Paul). – *Processus stochastiques et mouvement brownien.* – Gauthier-Villars, Paris, 1965.

[140] Louchard (Guy). – Kac's Formula, Lévy's Local Time and Brownian Excursion. *Journal of Applied Probability*, vol. 21, 1984, pp. 479–499.

[141] Louchard (Guy). – The Brownian Excursion Area : a Numerical Analysis. *Computers and Mathematics with Applications*, vol. 10, 1984, pp. 413–417.

[142] Louchard (Guy). – Erratum. *Computers and Mathematics with Applications A*, vol. 12, 1986, p. 375.

[143] Louchard (Guy) and Turner (John W.). – Generalized covariances of multi-dimensional Brownian Excursion local times. *Theoretical Computer Science*, vol. 21, 2003.

[144] Lyon (Gordon). – Syntax-directed least-errors analysis for context-free languages : A pratical approach. *Comm. of the ACM*, vol. 17, 1974, pp. 5–37.

[145] Mahmoud (Hosam). – On the distribution of leaves in rooted subtrees of recursive trees. *Annals of Applied Probability*, vol. 1, 1991, pp. 406–418.

[146] Mahmoud (Hosam). – *Evolution of Random Search Trees.* – John Wiley & Sons, Inc., 1992, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, xii+324p.

[147] Mahmoud (Hosam). – The joint distribution of the three types of nodes in uniform binary trees. *Algorithmica*, vol. 13, 1995, pp. 313–323.

[148] Mahmoud (Hosam) and Pittel (Boris). – On the most probable shape of a search tree grown from random permutations. *SIAM Journal on Algebraic and Discrete Methods*, vol. 1, 1984, pp. 69–81.

[149] Mahmoud (Hosam), Smythe (R.), and Szymanski (J.). – On the structure of plane-oriented recursive trees and their branches. *Random Structures and Algorithms*, vol. 4, 1993, pp. 151–176.

[150] Mahmoud (Hosam) and Smythe (Robert). – Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures and Algorithms*, vol. 3, 1988, pp. 255–266.

[151] Marckert (Jean-François). – The contour of size n general planar trees. `http://fermat.math.uvsq.fr/~marckert/contour.ps.gz`, 2000.

[152] Marckert (Jean-François). – Cours sur les arbres. `http://fermat.math.uvsq.fr/~marckert/tunis.ps.gz`, 2001.

[153] Marckert (Jean-François) and Mokkadem (Abdelkader). – Ladder variables, Internal structure of Galton-Watson trees and Finite branching random walks. *Journal of Applied Probability*, vol. 40, n°3, 2003, pp. 671–689.

[154] Martínez (Conrado). – *Statistics under the BST Model.* – Doctorat, Université Polytechnique de Catalogne, March 1992.

[155] Martínez (Conrado). – Partial quicksort. In *Proc. of the First ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO).* – 2004. Accepted.

[156] Martínez (Conrado), Panholzer (Alois), and Prodinger (Helmut). – On the number of descendants and ascendants in random search trees. *Electronic Journal of Combinatorics*, vol. 5, n#R20, 1998.

[157] Martínez (Conrado) and Roura (Salvador). – Randomized binary search trees. *J. ACM*, vol. 45, n°2, 1998, pp. 288–323.

[158] Martínez (Conrado) and Roura (Salvador). – Optimal sampling strategies in quicksort and quickselect. *SIAM J. on Computing*, vol. 31, n°3, 2001, pp. 683–705.

[159] McDiarmid (Colin J. H.) and Hayward (Ryan B.). – Large deviations for Quicksort. *Journal of Algorithms*, vol. 21, 1996, pp. 476–507.

[160] Mishkov (Rumen L.). – Generalisation of the formula of Faa di Bruno for a composite function with a vector argument. *Internat. J. Math. & Math. Sci.*, vol. 24, n°, 1999, pp. 481–491.

[161] Nebel (Markus E.). – On the Horton-Strahler number for combinatorial tries. *RAIRO Theoret Inf Appl*, vol. 34, 2000, pp. 279–296.

[162] Nebel (Markus E.). – A unified approach to the analysis of Horton-Strahler parameters of binary tree structures. *Random Structures and Algorithms*, vol. 21, n°3-4, 2002, pp. 252–277.

[163] Nebel (Markus E.). – *On a Statistical Filter for RNA Secondary Structures.* – Technical Report n°5/2002, Frankfurter Informatik-Berichte, 2002.

[164] Nebel (Markus E.). – The stack-size of combinatorial tries revisited. *Discrete Math Theoret Comput Sci*, vol. 5, 2002, pp. 1–16.

[165] Nebel (Markus E.). – The stack-size of tries : a combinatorial study. *Theoretical Computer Science*, vol. 270, n°1-2, 2002, pp. 441–461.

[166] Nebel (Markus E.). – Combinatorial Properties of RNA Secondary Structures. *Journal of Computational Biology*, vol. 9, n°3, 2003, pp. 541–574.

[167] Nebel (Markus E.). – *Identifying Good Predictions of RNA Secondary Structure.* – Technical Report n°5/2003, Frankfurter Informatik-Berichte, 2003.

[168] Nebel (Markus E.). – Investigation of the Bernoulli-Model of RNA Secondary Structures. *Bulletin of Mathematical Biology*, 2004. – To appear.

[169] Neininger (Ralph) and Rüschendorf (Ludger). – A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, 2002.

[170] Nguyễn Thế (Michel). – Enumeration of Geometric Configurations on a Convex Polygon, summary of a talk by Marc Noy. In Chyzak (Frédéric) (editor), *Algorithms seminar, 1999-2000.* pp. 17–22. – Research Report n°4056, INRIA, 2000.

[171] Nguyễn Thế (Michel). – Distribution of the Size of Simplified or Reduced Trees. In *Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities. Trends in Mathematics*, pp. 341–354. – Birkhäuser, Versailles, 2002.

[172] Nguyễn Thế (Michel). – Everything You Always Wanted to Know about Quicksort, but Were Afraid to Ask, summary of a talk by Marianne Durand. In Chyzak (Frédéric) (editor), *Algorithms seminar, 2001-2002.* pp. 57–62. – Research Report n°5003, INRIA, 2002.

[173] Nguyễn Thế (Michel). – Area of Brownian Motion with Generatingfunctionology. In Banderier (Cyril) and Krattenthaler (Christian) (editors), *Discrete Random Walks : Theory and Applications in Combinatorics, Computational Biology, Computer Science, Probabilities, and Statistical Physics.* pp. 229–242. – Discrete Mathematics and Theoretical Computer Science, Paris, Institut Henri Poincaré, 2003.

[174] Nguyễn Thế (Michel). – Area and Inertial Moment of Dyck Paths. *Combinatorics, Probability and Computing*, vol. 13, 2004. – To appear.

[175] Otter (Richard). – The number of trees. *Annals of Mathematics*, vol. 49, 1948, pp. 583–599.

[176] Perman (Mihael) and Wellner (Jon). – On the distribution of Brownian areas. *Annals of Applied Probability*, vol. 6, 1995, pp. 1091–1111.

[177] Petazzoni (Bruno). – *Seize problèmes d'informatique.* – Springer, Paris, 2001.

[178] Rachev (Svetlozar T.). – *Probability Metrics and the Stability of Stochastic Models.* – John Wiley, New York, 1991.

[179] Rachev (Svetlozar T.) and Rüschendorf (Ludger). – Propagation of chaos and contraction of stochastic mappings. *Siberian Advances in Mathematics*, vol. 1, 1994, pp. 114–150.

[180] Rachev (Svetlozar T.) and Rüschendorf (Ludger). – Probability metrics and recursive algorithms. *Advances in Applied Probability*, vol. 27, 1995, pp. 770–799.

[181] Régnier (Mireille). – A limiting distribution for quicksort. *RAIRO Theoretical Informatics and Applications*, vol. 23, 1989, pp. 335–343.

[182] Rémy (Jean-Luc). – Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Theoretical Informatics and Applications*, vol. 19, n°2, 1985, pp. 179–195.

[183] Revuz (Daniel) and Yor (Marc). – *Continuous Martingales and Brownian Motion, 3rd edition.* – Springer-Verlag, New York, 1998.

[184] Rivas (Elena) and Eddy (Sean R.). – A dynamic programming algorithm for RNA structure prediction including pseudoknots. *Journal of Molecular Biology*, vol. 285, 1999, pp. 2053–2068.

[185] Rogers (L. Chris G.) and Williams (David). – *Diffusions, Markov Processes, and Martingales.* – Cambridge University Press, 2000.

[186] Samorodnitsky (Gennady) and Taqqu (Murad S.). – *Stable non-gaussian random processes : stochastic models with infinite variance.* – Chapman and Hall, 1994.

[187] Schachinger (Werner). – Asymptotic normality of recursive algorithms via martingale difference arrays. *Discrete Mathematics and Theoretical Computer Science*, vol. 4, n°2, 2001, pp. 363–398.

[188] Sedgewick (Robert) and Flajolet (Philippe). – *Introduction à l'analyse des algorithmes.* – International Thomson Publishing, France, 1996. 492 pages. A translation of the original English version (ISBN 2-84180-957-9).

[189] Sedgewick (Robert) and Flajolet (Philippe). – *An Introduction to the Analysis of Algorithms.* – Addison-Wesley Publishing Company, 1996.

[190] Shepp (Lawrence A.). – On the Integral of the Absolute Value of the Pinned Wiener Process. *The Annals of Probability*, vol. 10, 1982, pp. 234–239.

[191] Smythe (Robert T.) and Mahmoud (Hosam M.). – A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, vol. 51, 1995, pp. 1–27.

[192] Soria-Cousineau (Michèle). – *Méthodes d'analyse pour les constructions combinatoires et les algorithmes.* – PhD thesis, Université de Paris-Sud, Orsay, July 1990.

[193] Stanley (Richard P.). – *Enumerative Combinatorics.* – Cambridge University Press, New York, 1999, vol. 2.

[194] Steyaert (Jean-Marc). – *Structure et complexité des algorithmes.* – Doctorat d'État, Université Paris VII, April 1984.

[195] Szpankowski (Wojciech). – *Average Case Analysis of Algorithms on Sequences.* – John Wiley & Sons, 2001.

[196] Tahi (Fariza), Engelen (Stefan), and Régnier (Mireille). – A fast algorithm for RNA secondary structure prediction including pseudoknots. In *3rd IEEE Symposium on Bioinformatics and Bioengineering (BIBE 2003). Bioinformatics*, pp. 11–17. – IEEE Computer Society proceedings, Bethesda, Maryland, 2003. March 10-12 2003.

[197] Tahi (Fariza), Gouy (Manolo), and Régnier (Mireille). – Automatic RNA secondary structure prediction with a comparative approach. *Computers and Chemistry*, vol. 26, n°5, 2002, pp. 521–530.

[198] Takács (Lajos). – A bernoulli excursion and its various applications. *Advances of Applied Probability*, vol. 23, 1991, pp. 557–585.

[199] Takács (Lajos). – On a probability problem connected with railway traffic. *Journal of Applied Mathematics and Stochastic Analysis*, vol. 4, n°1, 1991, pp. 1–27.

[200] Takács (Lajos). – On the distribution of the integral of the absolute value of the Brownian motion. *Annals of Applied Probability*, vol. 3, 1993, pp. 186–197.

[201] Tan (Kok Hooi) and Hadjicostas (Petros). – Some properties of a limiting distribution in Quicksort. *Statist. Probab. Lett.*, vol. 25, 1995, pp. 87–94.

[202] Tolmatz (Leonid). – Asymptotics of the distribution of the integral of the absolute value of the Brownian bridge for large arguments. *The Annals of Probability*, vol. 28, 2000, pp. 132–139.

[203] Tolmatz (Leonid). – Asymptotics of the distribution of the integral of the positive part of the Brownian bridge for large arguments. 2001. – `http ://www.tolmatz.net/PAPERS/04POSBB/30s.pdf`.

[204] Tolmatz (Leonid). – On the distribution of the square integral of the Brownian bridge. *The Annals of Probability*, vol. 30, 2002, pp. 253–269.

[205] Tolmatz (Leonid). – Addenda : On the distribution of the square integral of the Brownian bridge. *The Annals of Probability*, vol. 31, 2003, pp. 530–532.

[206] Tolmatz (Leonid). – The saddle point method for the integral of the absolute value of the Brownian motion. In Banderier (Cyril) and Krattenthaler (Christian) (editors), *Discrete Random Walks : Theory and Applications in Combinatorics, Computational Biology, Computer Science, Probabilities, and Statistical Physics*. pp. 309–324. – Discrete Mathematics and Theoretical Computer Science, Paris, Institut Henri Poincaré, 2003.

[207] Waldispühl (Jérôme), Behzadi (Behshad), and Steyaert (Jean-Marc). – An approximate matching algorithm for finding (sub-)optimal sequences in s-attributed grammars. In *Proceedings of the first European Conference on Computational Biology, ECCB 2002. Bioinformatics*, vol. 18, pp. 250–259. – OXFORD University Press, 2002.

[208] Waterman (Michael S.). – *Introduction to Computational Biology. Maps, Sequences and Genomes.* – Chapman & Hall, 1995.

[209] Waterman (Michael S.) and Smith (Temple F.). – RNA secondary structure : a complete mathematical analysis. *Math Biosci.*, vol. 42, 1978, pp. 257–266.

[210] Williams (David). – *Probability with Martingales.* – Cambridge University Press, 1991.

[211] Wong (Roderick). – *Asymptotic Approximations of Integrals.* – Academic Press, 1989.

[212] Xiong (Momiao) and Waterman (Michael S.). – A Phase Transition for the Minimum Free Energy of Secondary Structures of a random RNA. *Advances in Applied Mathematics*, vol. 18, 1997, pp. 111–132.

[213] Yor (Marc). – *Some Aspects of Brownian Motion, Part I : Some Special Functionals.* – Birkhäuser Verlag, Basel, Boston, Berlin, 1992, *Lectures in Mathematics.*

[214] Yor (Marc). – *Local times and excursions for brownian motion : a concise introduction.* – Postgrado de Matemáticas, Facultad de Ciencias, Universidad Central de Venezuala, Caracas, 1995. Lecciones en Mathemáticas, vol. 1.

[215] Zolotarev (Vladimir M.). – Approximation of the distributions of sums of independant random variables with values in infinite-dimensional spaces. *Theor. Probability Appl.*, vol. 21, 1976, pp. 721–737.

[216] Zucker (Michael) and Sankoff (David). – RNA secondary structure and their prediction. *Bull. Math. Biol.*, vol. 46, 1984, pp. 591–621.

Résumé de la thèse

Distribution de valuations sur les arbres

Résumé : Les arbres sont des structures combinatoires qui peuvent représenter plusieurs choses en informatique, comme des structures de données, des expressions booléennes ou arithmétiques, des arbres de dérivation en langages formels. Cette thèse étudie la distribution limite de paramètres récursifs (appelés "valuations") définis sur des arbres, lorsque la taille des arbres tend vers l'infini.

Dans la première partie, nous étudions des paramètres associés à des expressions arithmétiques. Nous étudions la distribution limite de résultats d'expressions arithmétiques tirées au hasard, avec différents jeux d'opérateurs, qui sont $\{\min, +\}$, $\{+, -\}$ et $\{+, \times\}$, et obtenons diverses loi limites. Nous étudions la loi limite de la taille d'arbres réduits lorsque nous simplifions des expressions booléennes ou avec des règles telles que (X ET $X \Longrightarrow X$), ($x - x = 0$), (X OU TRUE $\Rightarrow$ TRUE), etc. Nous étudions aussi des structures secondaires de molécules d'ARN, à travers des arbres de dérivation, ce qui constitue le premier pas dans la tentative de trouver un algorithme plus rapide pour déterminer la structure secondaire optimale d'un mot d'ARN donné.

Dans la seconde partie, nous étudions la distribution jointe de la longueur de cheminement et du moment d'inertie d'arbres et de divers chemins aléatoires. Nous étudions également la distribution jointe de la taille et de la longueur de cheminement interne de tries conditionnés par leur nombre de clés avec des sources de Bernoulli, et obtenons une loi limite gaussienne.

Distribution of Valuations on Trees

Abstract : Trees are combinatorial structures that can represent many things in computer science such as data structures, boolean or arithmetic expressions, derivation trees in formal languages. This thesis studies the limit distribution of recursive parameters (called "valuations") defined on trees, when the size of the trees tends to infinity.

In the first part, we study parameters associated with arithmetic expressions. We study the limit distribution of the results of arithmetic expressions drawn at random, with different sets of operators, that are $\{\min, +\}$, $\{+, -\}$ and $\{+, \times\}$, and obtain various limit laws. We study the limit law of the size of reduced trees when we simplify boolean or arithmetic expressions with rules such as (X AND $X \Longrightarrow X$), ($x - x = 0$), (X OR TRUE $\Rightarrow$ TRUE), and so on. We also study secondary structures of RNA molecules through derivation trees, which is the first step of an attempt to find a faster algorithm to compute the optimal secondary structure of a given RNA word.

In the second part, we study the joint distribution of the internal path length and the inertial moment of trees and of various random walks. We also study the joint distribution of the size and the internal path length of tries conditioned by number of keys with a Bernoulli source, and find a limit Gaussian distribution.

Laboratoire d'informatique École Polytechnique
91128 Palaiseau Cedex - FRANCE

(phone) 33 1 69 33 40 73 - Fax 33 1 69 33 30 14
lix@polytechnique.fr - http://lix.polytechnique.fr