THÈSE

présentée devant

L'UNIVERSITÉ DE VERSAILLES-SAINT-QUENTIN

pour obtenir le titre de :

DOCTEUR DE L'UNIVERSITÉ DE VERSAILLES-SAINT-QUENTIN

Mention : Mathématiques et Applications

PAR

THIERRY KLEIN

TITRE DE LA THÈSE :

INÉGALITÉS DE CONCENTRATION, MARTINGALES ET ARBRES ALÉATOIRES

Rapporteurs

M. Svante JANSON
M. LiMing WU

Soutenue publiquement le 19 décembre 2003 devant le jury composé de :

M. Jean BRETAGNOLLE	Examinateur
M. Oleksiy KHORUNZHIY	Examinateur
M. Michel LEDOUX	Examinateur
M. Emmanuel RIO	Directeur de thèse
M. LiMing WU	Rapporteur

Résumé

Inégalités de concentration, martingales et arbres aléatoires

Cette thèse comporte trois parties. Dans les deux premières parties, nous nous intéressons à deux aspects de la concentration de la mesure. Dans la dernière partie, nous nous intéressons à l'analyse asymptotique des arbres binaires de recherche dans le modèle dit des permutations aléatoires.

Dans la première partie, nous donnons des constantes optimales dans les inégalités de concentration de type Talagrand pour le maxima de processus empiriques associés à des variables aléatoires indépendantes. Notre méthode est basée sur la méthode dite de Herbst et sur des techniques de calculs entropiques.

Dans la seconde partie, nous prouvons des inégalités de concentration convexe pour des processus de comptage à temps discret et à temps continu. Nous appliquons ensuite ces résultats pour prouver que le sup de variables aléatoires indépendantes binomiales (resp. de Poisson) vérifie une inégalité de concentration convexe.

Dans la troisième partie, nous étudions le comportement asymptotique des arbres binaires de recherche. Nous utilisons essentiellement deux méthodes : le plongement et le tiltage. Nous obtenons alors de nouveaux résultats sur l'arbre binaire de recherche ainsi que des nouvelles preuves de résultats connus.

Abstract

Concentration inequlalities, martingales and random trees

This Phd thesis is divided into three parts. The two first parts deal with two different aspects of the concentration of the measure. In the third part, we are interested in the asymptotic analysis of the binary search tree under the random permutation model

In the first part, we give optimal constants in Talagrand's concentration inequalities for maxima of empirical processes associated to independent and eventually non identically distributed random variables. Our method is based on the so-called Herbst method.

In the second part, we prove convex concentration inequalities for discrete and continuous time counting processes. Then we apply these inequalities to prove that the supremum of independent binomial random variables and the supremun of independent Poisson random variables satisfy convex concentration inequalities.

In the third part, we are interested in the asymptotic analysis of the binary search tree under the random permutation model. Two methods are mainly used : the first one is the embedding in continuous time and the second one is the tilting probability method. We get new results on the binary search tree and also new proofs of known results.

"Il n'est de problèmes qu'un manque de solutions ne finisse par résoudre"

Henri Queuille

Remerciements

"C'est impossible..."
"C'est impossible de faire ce calcul..."

Combien de fois ai-je pensé cela pendant ma thèse ?
On s'approche certainement de l'infini !

Je tiens à remercier chaleureusement Emmanuel de m'avoir permis de me rendre compte de mon erreur, **c'était possible ! !**
En côtoyant Emmanuel, j'ai appris que l'on pouvait mener à bout des calculs difficiles. J'ai beaucoup apprécié son honnêteté intellectuelle, sa grande rigueur scientifique et la manière dont il a dirigé ma thèse. Je tiens à le remercier pour son aide précieuse, sa grande connaissance des mathématiques qu'il a su partager avec moi, ainsi que la confiance qu'il m'a accordée durant ces années de thèse. Ce travail lui doit beaucoup.

Un grand merci aux rapporteurs Svante Janson et Liming Wu d'avoir eu la gentillesse et la patience de lire en détails mes travaux.

Merci à Jean Bretagnolle, Oleksiy Khorunzhiy, Michel Ledoux, Emmanuel Rio et Liming Wu de me faire l'honneur de faire partie de mon jury.

Alain, Brigitte, Emmanuel et Jean-François (mes coauteurs), j'ai eu beaucoup de plaisir à travailler à vos côtés et je suis très fier que nous ayons à maintes reprises "inventé l'eau tiède et le fil à couper le beurre".

Jean-François mérite une médaille pour toute son aide. Merci pour ton aide mathématique, informatique, les pauses cafés, les bières, les parties de cartes...

Merci à toute l'équipe de Proba-Stats, en particulier à Alain qui, parce qu'il ne sait rien, a su répondre à toutes mes questions.

Merci à tous les membres du LAMA et du département de mathématiques pour votre accueil chaleureux pendant ces années de thèse.

Un petit coucou aux thésards, courage, c'est bientôt fini ! !
Un grand coucou à Laurence, merci d'avoir partagé le stress des derniers mois, ainsi que les nombreuses pauses pepitos.

Voici maintenant venu le temps de dire merci à toutes les personnes qui comptent beaucoup pour moi (que celles que je vais oublier me le pardonnent).

A ma famille, que je retrouve toujours avec bonheur qu'elle soit ici ou de l'autre côté de la terre.

A tous mes amis, je ne conçois pas la vie sans vous.

A ceux qui ont connu le chinois à 36, son côte de Provence et ses sakés.

A mes compagnons qui ont gouté aux joies des folles nuits du Paris médiéval, du mélo coton aux courses dans les escaliers (comprenne qui pourra!!).

A tous les lyonnais, amis de longues dates (certains depuis près de 15 ans), ceux qui se souviennent du Parc, des parties interminables de coinche (règle numéro 1 on ne coinche pas à 80), des jeudis de beaujolais nouveaux (faire attention aux radiateurs), des troisièmes mi-temps à la gauloise et de la Panda qui m'a si souvent ramenée chez moi.

A tous ceux de la cité, ceux qui se souviennent des bars et des soirées de la Deutsh et ceux qui m'ont fait découvrir tant de cultures différentes.

A tous ceux, qui ne rentrent dans aucunes de ces catégories, que j'ai rencontrés sur un terrain de rugby, ou qui depuis ma rencontre se méfient des cotons tiges (encore une fois comprenne qui pourra!!).

A tous les idéalistes et les rêveurs...

Enfin, à Marta et à mon futur fils.

A vous tous,

je vous dédie ce travail, vous en êtes tous un peu responsable.

Table des matières

Première partie

Introduction

L'objectif initial de ce travail était d'établir des inégalités de concentration pour des fonctionnelles de variables aléatoires (éventuellement liées aux graphes aléatoires) indépendantes. Deux axes principaux étaient envisagés :

• Essayer d'obtenir les constantes optimales (conjecturées par Massart [38]) dans les inégalités de concentration de type Talagrand pour le supremum de fonctionnelles de variables aléatoires indépendantes.

• Essayer d'obtenir des inégalités de concentration convexe (en particulier pour le problème classique $\sup\{\sum_{i=1}^n f(\xi_i) : f \in \mathcal{F}\}$). Ces dernières étant plus précises que les inégalités de concentration de type Talagrand.

L'idée était d'appliquer ensuite ces inégalités de concentration à des structures combinatoires comme la plus grande sous-suite croissante dans une permutation aléatoire, le nombre de cliques ou le nombre chromatique dans un graphe aléatoire ou encore la hauteur d'un arbre binaire de recherche afin d'obtenir des bornes non asymptotiques pour ces quantités.

Nous avons réussi à réaliser partiellement ces objectifs. Nous avons obtenu pour les inégalités de concentration de type Talagrand des constantes proches des constantes optimales en particulier avec un facteur variance asymptotiquement exact. Nous avons établi des inégalités de concentration convexe pour des processus de sauts purs (en particulier pour des variables négativement associées). Dans le cas des parties indépendantes ces inégalités s'appliquent et donnent une inégalité de concentration convexe pour $\sup\{\mathbb{P}_n(A); A \in \mathcal{A}\}$.

Ensuite nous avons essayé d'appliquer ces théorèmes de concentration convexes à la plus grande sous-suite croissante dans une permutation aléatoire. Il est possible de vérifier (à la main) les hypothèses de ces théorèmes pour les petites valeurs de n ($n \leq 12$), mais leur vérification pour toutes les valeurs de n n'est pas immédiate. Afin de mieux comprendre ces phénomènes, nous avons décidé d'approfondir l'étude probabiliste de cette structure. Comme la plus grande sous-suite croissante dans une permutation aléatoire ressemble à certaines quantités relatives aux arbres binaires de recherche (comme par exemple la hauteur), nous avons donc étudié en détail la façon dont évolue les arbres binaires de recherche.

Les objets considérés

• Le maximum des processus empiriques associé à des variables aléatoires indépendantes :

On étudie les propriétés de concentration autour de leur moyenne des quantités Z de la forme

$$Z = \sup\{s^1(X_1) + \cdots + s^n(X_n), (s^1, \ldots, s^n) \in \mathcal{S}\}, \tag{1}$$

où $X_1, X_2, \ldots$ sont des variables indépendantes à valeurs dans un espace polonais $\mathcal{X}$, et les s^i sont des fonctions de $\mathcal{X}$ dans $[-1, 1]$.

Pour cette étude, nous utilisons des techniques entropiques (log-Sobolev) et la méthode dites de Herbst.

• Les processus dits de sauts purs à temps discret ou continu :

Il s'agit de processus constants par morceaux dont les sauts valent au plus 1. Nous étudions leurs propriétés de concentration pour les fonctions convexes. On utilise des techniques liées aux variables négativement associées, ainsi que des techniques de calcul stochastique pour les processus de sauts purs croissants.

• Les arbres binaires de recherche :

Il s'agit de l'étude d'un modèle probabiliste d'une structure combinatoire classique utilisée en informatique pour stocker de manière efficace des données. Nous étudions la forme assymptotique de cet arbre en utilisant des techniques de martingales et de processus de branchement. Les deux points clés de notre approche sont le plongement et le tiltage.

Cette thèse comporte donc trois parties. Dans la première partie nous donnons des constantes précises dans les inégalités de concentration de type Talagrand pour les processus empiriques indexés par des classes de parties. Dans la seconde nous donnons des inégalités de concentration convexe pour des processus de sauts purs. Finalement dans la troisième partie, nous nous intéressons à l'analyse asymptotique des arbres binaires de recherche dans le modèle dit des permutations aléatoires.

Première partie

Les inégalités de concentration de type Talagrand sont une extension aux fonctions de variables indépendantes des inégalités exponentielles de Bennett ou de Bernstein pour les sommes de variables indépendantes. Ce type d'inégalités est apparu la première fois dans les travaux de Hoeffding [24].

Théorème 1 (Hoeffding [24]) *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes telles que X_i soit à valeurs dans $[a_i, b_i]$. Soit $S_n = \sum_{i=1}^n \big(X_i - \mathbb{E}(X_i)\big)$, alors pour tout $x > 0$, on a*

$$\mathbb{P}(S_n \geq x) \leq \exp\Big(-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big). \tag{2}$$

Azuma [6] a étendu (2) aux martingales à accroissements bornés puis McDiarmid [35, 36] a montré comment appliquer les inégalités d'Azuma pour obtenir des inégalités de concentration pour les fonctionnelles 1-lipshitziennes en distance de Hamming de variables aléatoires indépendantes (pour plus de détails le lecteur pourra regarder l'ouvrage de Ledoux [31] et le cours de Saint-Flour de Massart [39], qui donnent un résumé des connaissances actuelles sur les phénomènes de concentration de la mesure).

Les inégalités exponentielles de type Hoeffding ne prennent pas en compte la variance de S_n. Elles sont fondées sur des majorations de la transformée de Laplace de S_n. Pour améliorer ces inégalités (dans les bandes de moyennes déviations), les majorations ultérieures de la transformée de Laplace de S_n ont pris en compte la variance de S_n. Les résultats obtenus ainsi sont les inégalités dites de Bennett [8] et de Bernstein.

Théorème 2 (Bennett) *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de carré intégrable telles qu'il existe une constante positive b, pour laquelle $X_i \leq b$ pour tout $i \leq n$. Soit $S_n = \sum_{i=1}^n \big(X_i - \mathbb{E}(X_i)\big)$ et $v = \sum_{i=1}^n \mathbb{E}(X_i^2)$. Alors pour tout $x > 0$, on a*

$$\mathbb{P}(S_n \geq x) \leq \exp\Big(-\frac{v}{b^2} h\big(\frac{bx}{v}\big)\Big), \tag{3}$$

où $h(u) = (1+u)\log(1+u) - u$ pour $u > -1$.

Pour les variables non bornées (ayant cependant une transformée de Laplace finie sur un voisinage de 0), il est encore possible d'obtenir une inégalité de type Bernstein.

Théorème 3 (Inégalité de type Bernstein) *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes. Supposons qu'il existe des nombres strictement positifs v et c tel que pour tout entier $k \geq 2$*

$$\sum_{i=1}^{n} \mathbb{E}\left(|X_i|^k\right) \leq \frac{k!}{2} v c^{k-2}. \tag{4}$$

Soit $S_n = \sum_{i=1}^{n} X_i - \mathbb{E}(X_i)$, alors pour tout x strictement positif on a

$$\psi^{\star}_{S_n}(x) \geq \frac{v}{c^2} h_1\Big(\frac{cx}{v}\Big) \tag{5}$$

où $h_1(u) = 1 + u - \sqrt{1+2u}$, pour $u > 0$, *et $\psi^{\star}_{S_n}$ est la transformée de Cramer de la log-Laplace de S_n.*
En particulier, pour tout x strictement positif

$$\mathbb{P}(S_n \geq \sqrt{2vx} + cx) \leq \exp(-x). \tag{6}$$

Le résultat ci-dessus est dû à Birgé et Massart [14]. La forme de l'équation (6) provient du fait que $h_1^{-1}(x) = x + \sqrt{2x}$ (voir par exemple Rio [46]).

Talagrand ([49], théorème 4.2.) donne une extension du résultat de Bennett aux fonctionnelles de variables aléatoires indépendantes. Les inégalités de Talagrand s'appliquent en particulier à la fonctionnelle $Z_n = \sup\{S_n(f),\ f \in \mathcal{F}\}$, où $S_n(f) = f(X_1) + \cdots + f(X_n)$. On obtient alors :

Théorème 4 (Talagrand) *Soient $\xi_1, \ldots, \xi_n$ des variables aléatoires indépendantes à valeurs dans un espace mesurable $(\mathbb{X}, \mathcal{X})$. Soit $\mathcal{F}$ une famille dénombrable de fonctions mesurables à valeurs dans $\mathbb{R}$, vérifiant $\|f\|_\infty \leq b < \infty$ pour tout $f \in \mathcal{F}$. Soient $Z = \sup\{\sum_{i=1}^{n} f(\xi_i) : f \in \mathcal{F}\}$ et $v = \mathbb{E}\Big(\sup\{\sum_{i=1}^{n} f^2(\xi_i) : f \in \mathcal{F}\}\Big)$. Alors pour tout x strictement positif,*

$$\mathbb{P}\Big(Z \geq \mathbb{E}(Z) + x\Big) \leq K \exp\Big(-\frac{1}{K'}\frac{x}{b}\log\big(1 + \frac{xb}{v}\big)\Big) \tag{7}$$

et

$$\mathbb{P}\Big(Z \geq \mathbb{E}(Z) + x\Big) \leq K \exp\Big(-\frac{x^2}{2(c_1 v + c_2 b x)}\Big), \tag{8}$$

où K, K', c_1 et c_2 sont des constantes universelles strictement positives. De plus, les mêmes inégalités sont valables si l'on remplace Z par $-Z$.

La preuve de Talagrand repose sur des inégalités isopérimétriques pour les mesures produits. Ledoux [30] retrouve ces inégalités par des méthodes entropiques en utilisant la méthode dite de Herbst. Il obtient (8) avec $K = 2$, $c_1 = 42$ et $c_2 = 8$. Mais il prend v de la forme $\mathbb{E}\big(\sup\{\sum_{i=1}^{n} f^2(\xi_i) : f \in \mathcal{F}\}\big) + Cb\mathbb{E}(Z)$. Les techniques de preuves que nous utilisons reposent essentiellement sur la méthode de Herbst. C'est pourquoi nous commençons par exposer cette méthode en quelques mots. On se donne $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires indépendantes à valeurs dans un espace polonais $\mathcal{X}$ et $\mathcal{F}$ une classe dénombrable de fonctions mesurables. On note

$$S_n(f) = \sum_{k=1}^{n} f(X_i)$$

et on considère la quantité

$$Z_n = \sup_{f\in\mathcal{F}}\{S_n(f)\}. \tag{9}$$

L'objectif est de déterminer les "meilleures" fonctions possibles $f_d(x,n)$ et $f_g(x,n)$, (aussi proche que possible des fonctions que l'on obtient dans le cas des variables gaussiennes) pour lesquelles on ait pour tout $n \in \mathbb{N}$ et tout $x > 0$

$$\mathbb{P}\Big(Z_n - \mathbb{E}(Z_n) \geq x\Big) \leq f_d(x,n), \tag{10}$$

$$\mathbb{P}\Big(Z_n - \mathbb{E}(Z_n) \leq -x\Big) \leq f_g(x,n). \tag{11}$$

Explicitons la méthode de Herbst pour obtenir l'equation (10). Soit $t > 0$, alors

$$\mathbb{P}\Big(Z_n - \mathbb{E}(Z_n) \geq x\Big) = \mathbb{P}\Big(e^{tZ_n} \geq e^{tx+t\mathbb{E}(Z_n)}\Big).$$

En appliquant l'inégalité de Markov on a

$$\mathbb{P}\Big(Z_n - \mathbb{E}(Z_n) \geq x\Big) \leq e^{-tx-t\mathbb{E}(Z_n)+L_{Z_n}(t)} \tag{12}$$

où $L_{Z_n}(t) = \log\Big(\mathbb{E}\big(e^{tZ_n}\big)\Big)$ est la log-Laplace de Z_n. Ainsi un contrôle précis de L_{Z_n} donnera une fonction $f_d(x,n)$ convenable. Ledoux a montré comment une technique entropique permet d'obtenir un bon contrôle de L_{Z_n}. Le lemme suivant de tensorisation de l'entropie dû à Ledoux [30] (on retrouve aussi ce lemme sous des formes similaires dans les travaux de Massart [38] et de Rio [47]) s'avère être très utile pour la majoration de la log-Laplace de Z_n.

Lemme 1 (Ledoux) *Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes à valeurs dans un espace polonais $\mathcal{X}$. Soit $\mathcal{F}_n$ la tribu engendrée par $X_1,\ldots,X_n$ et $\mathcal{F}_n^k$ la tribu engendrée par $X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_n$. On note E_n^k l'opérateur d'espérance conditionnelle associé à $\mathcal{F}_n^k$ et on se donne une fonction f strictement positive $\mathcal{F}_n$ -mesurable, vérifiant $\mathbb{E}(f\log f) < \infty$. Alors*

$$\mathbb{E}(f\log f) - \mathbb{E}(f)\log\mathbb{E}(f) \leq \sum_{k=1}^{n} \mathbb{E}\big(f\log(f/E_n^k f)\big). \tag{13}$$

Si on applique ce lemme à $f(t) = e^{tZ_n}$ la partie gauche de l'inégalité (13) est égale à

$$tF'(t) - F(t)\log F(t)$$

où $F(t) = \mathbb{E}(\exp(tZ_n))$ est la transformée de Laplace de Z_n. Ainsi, si on arrive à contrôler la partie droite de (13) par un terme de la forme $F(t)V(t)$ (la fonction V faisant intervenir le terme de variance v), on obtient après division par $F(t)$

$$tL'(t) - L(t) \leq V(t). \tag{14}$$

Le lemme 1 fournit donc une inéquation différentielle vérifiée par L dont l'intégration fournit un contrôle de L (dépendant de v).

Massart [38] obtient des constantes plus précises que celles de Ledoux avec v de la forme $\sup\{\sum_{i=1}^n \mathbb{E}\big(f^2(\xi_i)\big) : f \in \mathcal{F}\} + Cb\mathbb{E}(Z)$, par des méthodes analogues à celles utilisées par Ledoux, et conjecture que les inégalités (7) et (8) sont valables avec $K = 1$, $K' = 1$, $c_1 = 1$ et $c_2 = 1/3$. Ensuite Rio [47] obtient des inégalités de Talagrand avec un facteur de variance asymptotiquement exact et Bousquet [15] et [16] répond positivement à la conjecture de Massart pour la concentration à droite lorsque les variables aléatoires $(X_1, \ldots, X_n)$ sont indépendantes et de même loi. Dans la suite, on appellera inégalité de concentration à droite une inégalité de concentration du type de (10), et une inégalité de concentration à gauche une inégalité de concentration du type de (11).

Dans la première partie de cette thèse, nous donnons des constantes précises pour les inégalités de Talagrand dans les deux cas suivants.

• Lorsque les variables aléatoires $(X_1, \ldots, X_n)$ sont indépendantes et de même loi, nous donnons l'inégalité de concentration à gauche suivante :

Théorème 5 *Soit $\mathcal{F}$ classe dénombrable de fonctions de $\mathcal{X}$ dans $]-\infty, 1]$, mesurables, de carré intégrable et d'espérance nulle sous $\mathbb{P}$. Soit σ^2 telle que $\sigma^2 \geq P(f^2)$ pour toute f dans $\mathcal{F}$.*
• Si les fonctions de $\mathcal{F}$ sont à valeurs dans $[-1, 1]$ alors pour tout $x > 0$,

$$\mathbb{P}\Big(Z \leq \mathbb{E}(Z) - x\Big) \leq \exp\Big(-\frac{v_n}{16}h(\frac{4x}{v_n})\Big), \tag{15}$$

avec $v_n = n\sigma^2 + 2\mathbb{E}(Z)$ et $h(x) = (1 + x)\log(1 + x) - x$.
• Si pour tout $f \in \mathcal{F}$, et pour tout $p \geq 2$, $\mid \mathbb{E}(f^p(X_i)) \mid \leq \frac{\sigma^2 p!}{2}$ alors

$$\forall t \in [0, 1[\qquad L(t) \leq -te^{-t}\mathbb{E}(Z) + \frac{n\sigma^2}{2}\frac{t^2(1 + 2t)}{1 - t}, \tag{16}$$

et par conséquent pour tout x positif,

$$\mathbb{P}\Big(Z \leq \mathbb{E}(Z) - x\Big) \leq \exp\Big(-(u - v)^2(1 - 2\frac{v(u - v)}{u^2})\Big), \tag{17}$$

où $L(t)$ est la log-*Laplace de $-Z$, $u = \sqrt{x + v_n/2}$, $v = \sqrt{v_n/2}$.*

On peut remarquer que les constantes obtenues sont moins bonnes que celles qu'obtient Bousquet ([15], [16]) dans son théorème de concentration à droite, cependant la deuxième partie fournit une inégalité de concentration sous des hypothèses plus faibles ; en effet on ne suppose plus les variables bornées mais seulement que leurs moments vérifient $\mid \mathbb{E}(f^p(X_i)) \mid \leq \frac{\sigma^2 p!}{2}$.

• Lorsque les variables aléatoires $(X_1, \ldots, X_n)$ sont indépendantes, nous obtenons (travail en commun avec Rio) les inégalités de concentration suivantes. La première est une inégalité de concentration à droite et la seconde une inégalité de concentration à gauche.

Théorème 6 *Soit $\mathcal{S}$ une classe dénombrable de fonctions mesurables à valeurs dans $[-1,1]^n$. On suppose que $\mathbb{E}(s^i(X_i)) = 0$ pour tout $s = (s^1, \ldots, s^n)$ dans $\mathcal{S}$ et tout entier i dans $\{1, \ldots, n\}$. Posons $V_n = \sup\{\text{var}\big(S_n(s)\big) : s \in \mathcal{S}\}$. On note L la log-Laplace de Z. Alors, pour tout $t > 0$,*

$$L(t) \leq t\mathbb{E}(Z) + \frac{t}{2}(2\mathbb{E}(Z) + V_n)(\exp(e^t - 1) - 1). \tag{18}$$

Par conséquent, pour tout $x > 0$,

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + x) \leq \exp\Big(-\frac{x}{2}\log\Big(1 + \log(1 + x/(2\mathbb{E}(Z) + V_n))\Big)\Big) \tag{19}$$

et

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + \sqrt{2x(V_n + 2\mathbb{E}(Z))} + x) \leq \exp(-x). \tag{20}$$

Théorème 7 *Sous les hypothèses du théorème 6, pour tout t dans $[0, t_0]$,*

$$L(-t) \leq -t\mathbb{E}(Z) + \frac{t}{9}(2\mathbb{E}(Z) + V_n)(\exp(3t) - 3t - 1). \tag{21}$$

Pour tout $x > 0$,

$$\mathbb{P}(Z \leq \mathbb{E}(Z) - x) \leq \exp\Big(-\frac{2\mathbb{E}(Z) + V_n}{9} h\Big(\frac{3x}{2\mathbb{E}(Z) + V_n}\Big)\Big), \tag{22}$$

où $h(x) = (1 + x)\log(1 + x) - x$.

La preuve du premier théorème que nous présentons (bien que plus complexe que les preuves existantes dans le cas i.i.d.) utilise les techniques log-Sobolev et de Herbst décrites ci-dessus. Pour prouver le second théorème, nous avons dû appliquer ces techniques non pas au processus Z, mais au processus empirique compensé Z_t défini par

$$Z_t := \sup\{S_n(s_i) + t^{-1}\log\mathbb{E}\Big(\exp(-tS_n(s_i))\Big) : 1 \leq i \leq m\}.$$

Il est à noter que les constantes obtenues dans le théorème 7 sont meilleures que celles obtenues dans le théorème 5. Cependant nous avons une hypothèse plus forte sur les queues de distribution à gauche des variables $s^i(X_i)$.

Seconde partie

Les inégalités de concentration ci-dessus sont déduites d'une majoration de $F_{Z-\mathbb{E}(Z)}(t) = \mathbb{E}\big(\phi_t(Z - \mathbb{E}(Z))\big)$ via l'inégalité de Markov, où ϕ_t est la fonction qui à x associe e^{tx}. Notons

$M(\phi_t, Z)$ le majorant de $F_{Z-\mathbb{E}(Z)}(t)$. En effet, l'application de Markov conduit à l'inégalité suivante :

$$\mathbb{P}\Big(Z - \mathbb{E}(Z) \geq x\Big) \leq \inf_{t>0} \left\{ \frac{M(\phi_t, Z)}{\phi_t(x)} \right\}. \tag{23}$$

La majoration recherchée est une majoration poissonnienne : on cherche à montrer que $\mathbb{E}\big(\phi_t(Z-\mathbb{E}(Z))\big) \leq \mathbb{E}\big(\phi_t(N-\mathbb{E}(N))\big)$, où N est une variable aléatoire de Poisson de paramètre adéquat.

Les fonctions ϕ_t considérées sont des fonctions convexes particulières. Ainsi, on obtiendra une amélioration des inégalités de concentration si l'infimum du membre de droite de (23) est remplacé par

$$\inf_{\phi \in \mathcal{C}} \left\{ \frac{M(\phi, Z)}{\phi(x)} \right\},$$

où $\mathcal{C}$ est l'ensemble des fonction convexes.

Ce constat motive l'introduction de la concentration convexe que l'on traite dans la deuxième partie de la thèse.

Dans la seconde partie, nous nous intéressons à des inégalités de concentration convexe. On dira qu'une variable aléatoire X est plus concentrée que Y pour les fonctions convexes si pour toute fonction convexe ϕ on a $\mathbb{E}\big(\phi(X)\big) \leq \mathbb{E}\big(\phi(Y)\big)$. Ce concept a été introduit par Hoeffding [24]. Il montre le résultat suivant.

Théorème 8 (Hoeffding) *Soient $b_1, \ldots, b_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre p_i et $S_n = b_1 + \cdots + b_n$. Posons $\bar{p} = \dfrac{p_1 + \ldots + p_n}{n}$ alors pour toute fonction convexe ϕ,*

$$\mathbb{E}\Big(\phi(S_n)\Big) \leq \mathbb{E}\Big(\phi\big(B(n, \bar{p})\big)\Big). \tag{24}$$

Dans la suite on dira qu'une variable aléatoire Z vérifie une inégalité de concentration convexe de type binômial si pour toute fonction convexe ϕ

$$\mathbb{E}\big(\phi(Z)\big) \leq \mathbb{E}\big(\phi(Y)\big), \tag{25}$$

où Y est une variable aléatoire de loi binômiale $\mathcal{B}(n, \mathbb{E}(Z)/n)$.

De même, on dira qu'une variable aléatoire Z vérifie une inégalité de concentration convexe de type Poisson, si pour toute fonction convexe ϕ

$$\mathbb{E}\big(\phi(Z)\big) \leq \mathbb{E}\big(\phi(Y)\big), \tag{26}$$

où Y est une variable aléatoire de loi de Poisson $\mathcal{P}(\mathbb{E}(Z))$. Hoeffding explique ensuite comment ce type d'inégalités permet de contrôler les queues de distribution. Notons tout de même que si on considère pour t strictement positif des fonctions convexes de la forme $\phi_t(x) = \exp(tx)$, et si Z vérifie une inégalité de concentration convexe de type binômial, on a, en appliquant l'inégalité de Markov

$$\mathbb{P}\Big(Z \geq \mathbb{E}(Z) + x\Big) \leq \exp\Big(L(t) - tx - t\mathbb{E}(Z)\Big), \tag{27}$$

où L est le logarithme de la transformée de la place d'une variable suivant la loi $\mathcal{B}(n, \mathbb{E}(Z)/n)$. Comme la fonction L est bien connue on peut optimiser le membre de droite de (27) suivant les valeurs de t et obtenir des majorations précises de la queue de distribution de Z. Il en va de même si la variable Z vérifie une inégalité de concentration convexe de type Poisson.

Bretagnolle [18] donne une version fonctionnelle du résultat d'Hoeffding, puis Pinelis ([40] et [41]) étudie un cas plus général où les fonctions ϕ appartiennent à une classe générale de fonctions. Shao [48] traite le cas des suites de variables aléatoires négativement associées (N.A). Comme on le verra dans le chapitre 3, le caractère N.A. est essentiel pour l'obtention de notre premier résultat de concentration convexe. C'est pourquoi nous rappelons ici brièvement une partie du travail de Shao. Rappelons qu'une suite finie de variables aléatoires $\{X_i,\ 1 \leq i \leq n\}$ est dite N.A si pour toute paire de sous-ensembles disjoints A_1 et A_2 de $\{1, 2, \ldots, n\}$

$$\mathrm{Cov}\{f_1(X_i,\ i \in A_1), f_2(X_j,\ j \in A_2)\} \leq 0, \tag{28}$$

si f_1 et f_2 sont des fonctions croissantes coordonnées par coordonnées et que les covariances existent. Shao montre le résultat suivant.

Théorème 9 (Shao) *Soit $\{X_i, 1 \leq i \leq n\}$ une suite de variables aléatoires négativement associée et soit $\{X_i^\star, 1 \leq i \leq n\}$ une suite de variables aléatoires indépendantes, pour laquelle X_i et $X_i^\star$ ont la même loi pour tout $i = 1, 2, \ldots, n$. Alors*

$$\mathbb{E}\Big(f\big(\sum_{i=1}^{n} X_i\big)\Big) \leq \mathbb{E}\Big(f\big(\sum_{i=1}^{n} X_i^\star\big)\Big) \tag{29}$$

pour toute fonction convexe f sur $\mathbb{R}$, dès que l'espérance du terme de droite de (29) existe.

Shao explique comment à partir de ce résultat on peut retrouver des inégalités classiques, comme l'inégalité maximale de Rosenthal ou l'inégalité de Kolmogorov. En particulier, il est capable d'étendre l'inégalité de Hoeffding pour une somme de variables aléatoires tirées sans remise dans une population finie. Pinelis [41] puis Bentkus [9] obtiennent des inégalités précises pour $\mathbb{P}(Z \geq x)$ à partir d'inégalités de concentration convexe.

Nous démontrons tout d'abord des inégalités de concentration convexe dans les deux cas suivants.

- Nous considérons d'abord des processus de comptage $(Z_n)_{n\in\mathbb{N}}$ à temps discret i.e.

$$Z_{n+1} - Z_n = 0 \text{ ou } Z_{n+1} - Z_n = 1. \tag{30}$$

On suppose que $(Z_n)_{n\in\mathbb{N}}$ vérifie l'hypothèse suivante

Hypothèse 1 *Pour tout n fixé, la suite $\Big(\mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k)\Big)_{k\geq 0}$, est décroissante.*

Alors on a le théorème suivant

Théorème 10 *Sous l'hypothèse 1, le processus $(Z_n)_{n\in\mathbb{N}}$ vérifie une inégalité de concentration convexe de type binômiale. En d'autres termes, pour toute fonction ϕ convexe*

$$\mathbb{E}(\phi(Z_n)) \leq \mathbb{E}(\phi(Y_n)), \tag{31}$$

où Y_n suit une loi $\mathcal{B}(n, \mathbb{E}(Z_n)/n)$.

Ce théorème est un corollaire du résultat de Shao car on verra que l'hypothèse 1 implique que les variables $(Z_n, Z_{n+1} - Z_n)$ sont N.A..

• Ensuite on considère des processus de comptage $(A_t)_{t\geq 0}$ à temps continu et on donne l'analogue du théorème 10. L'hypothèse 1 est remplacée par une hypothèse sur la dérivée (au sens de Lebesgue) du compensateur du processus $(A_t)_{t\geq 0}$. On rappelle que si $(A_t, \mathcal{F}_t)_{t\geq 0}$ est un processus de comptage alors il existe un unique processus croissant prévisible $(\Lambda_t)_{t\geq 0}$ tel que $(A_t - \Lambda_t)_{t\geq 0}$ soit une $\mathcal{F}_t$-martingale (pour une étude détaillée de ces processus nous renvoyons aux ouvrages de Brémaud [17] et de Dellacherie et Meyer [22]). Intuitivement, si on note λ_t la dérivée de Λ_t au sens de Lebesgue (l'hypothèse 2 assure l'existence d'une telle dérivée) alors λ_t représente la probabilité de sauts à l'instant t du processus $A_.$. Ainsi il est naturel de faire les deux hypothèses suivantes :

Hypothèse 2 *Le compensateur $(\Lambda_t)_{t\geq 0}$ est absolument continu et p.s. fini sur $[0,T]$ pour tout $T > 0$.*

Hypothèse 3 *$\mathbb{E}(\lambda_t \mid A_{t-})$ est une fonction décroissante de A_{t-}.*

L'hypothèse 2 implique que A a p.s. un nombre fini de sauts sur tout intervalle $[0,T]$. Nous obtenons alors le résultat suivant

Théorème 11 *Sous les hypothèses 2 et 3, le processus $(A_t)_{t\geq 0}$ vérifie une inégalité de concentration convexe de type Poisson. En d'autres termes, pour toute fonction convexe ϕ*

$$\mathbb{E}\big(\phi(A_t)\big) \leq \mathbb{E}\big(\phi(Y_t)\big), \tag{32}$$

où Y_t suit la loi $\mathcal{P}(\mathbb{E}(A_t))$.

La preuve que nous donnons de ce théorème utilise des techniques d'équations différentielles. L'idée est la suivante, on note $h(t) = \mathbb{E}\big(\phi(A_t)\big)$ et $g(t) = \mathbb{E}\big(\phi(Y_t)\big)$, on trouve ensuite une équation différentielle dont g est solution et dont h est sous-solution. Les deux points essentiels dans la preuve sont les suivants : tout d'abord on calcule explicitement le compensateur de $\big(\phi(A_t)-\phi(A_0)\big)_{t\geq 0}$, ensuite on note que $\big(\phi(1+A_{t-})-\phi(A_{t-}), \mathbb{E}(\lambda_t \mid A_{t-})\big)$ est négativement associé.

Une fois les théorèmes 10 et 11 démontrés, nous les appliquons respectivement au supremun de variables binômiales indépendantes et au supremum de variables de Poisson indépendantes. Dans le cas des variables binômiales, on construit un processus $(Z_u)_{u\in\mathbb{N}}$ vérifiant les hypothèses du théorème 10 et pour lequel à l'instant $u = nl$ on a $Z_{nl} = \sup\{S_n(p_1), \ldots, S_n(p_l)\}$ où les $S_n(p_i)$ sont indépendantes de loi $\mathcal{B}(n, p_i)$. En appliquant alors le théorème 10, on montre que Z_{nl} est plus concentré au sens des fonctions convexes qu'une variable Y de loi $\mathcal{B}(nl, \mathbb{E}(Z_{nl})/nl)$. De même dans le cas des variables de Poisson, on construit un processus $(Z_t)_{t\geq 0}$ vérifiant les hypothèses du théorème 11 et pour lequel à l'instant $t = 1$ on a $Z_1 = \sup\{N^{(1)}, \ldots, N^{(l)}\}$ où les $N^{(i)}$ sont indépendantes de loi $\mathcal{P}(\mu_i)$. En appliquant alors le théorème 11, on montre que Z_1 est plus concentré au sens des fonctions convexes qu'une variable Y de loi $\mathcal{P}(\mathbb{E}(Z_1))$.

Dans la fin de la deuxième partie (Chapitre 4), nous généralisons le théorème 11 au cas de processus $(A_t)_{t\geq 0}$ de sauts purs. C'est-à-dire un processus constant par morceaux dont les sauts sont bornés par 1. On prend $A_t = \int_0^t J_{s-} dB_s$ où $(B_s)_{s\geq 0}$ est un processus ponctuel dont le compensateur est $\Lambda_s = \int_0^s \lambda_u du$ et où $(J_s)_{s\geq 0}$ est un processus vérifiant $0 \leq J_s \leq 1$. On obtient alors un résultat ce concentration convexe faisant intervenir le compensateur C_t de A. Ce compensateur est défini par $C_t = \int_0^t \lambda_u J_{u-} du$. Cette inégalité de concentration convexe compare la martingale $A_t - C_t$ avec une variable de Poisson recentrée.

Théorème 12 *Soit $(A_s)_{s\geq 0}$ un processus de sauts purs. Supposons qu'il existe une constante $C > 0$ tel que $\|C_t\|_\infty \leq C$. Alors, pour toute fonction convexe ϕ, tel qu'il existe un $A > 0$ pour lequel $\lim_{x\to+\infty} \phi(x)\exp(-Ax) = 0$:*
a) La fonction

$$g(s) = \mathbb{E}\Big(\phi\big(A_s - N_{C_s} + N_C - C\big)\Big) \tag{33}$$

est décroissante.
b) Par conséquent

$$\mathbb{E}\Big(\phi(A_t - C_t)\Big) \leq \mathbb{E}\Big(\phi(N_C - C)\Big). \tag{34}$$

Les techniques de preuves sont similaires à celles utilisées dans le chapitre 3. Remarquons que dans l'équation (34) la variance de $N_C - C$ peut être bien supérieure à la variance de $A_t - C_t$, c'est pourquoi nous donnons par la suite deux inégalités de concentration convexe (théorèmes 13 et 14) faisant intervenir le compensateur quadratique $D_t := \int_0^t \lambda_u J_{u-}^2 du$ de $A_t - C_t$.

Théorème 13 *Soit $(A_s)_{s\geq 0}$ un processus de sauts purs. Supposons qu'il existe une constante $D > 0$ tel que $\|D_t\|_\infty \leq D$. Alors pour toute fonction convexe ϕ de classe C^2 dont la dérivée seconde est croissante et tel qu'il existe un $A > 0$ pour lequel $\lim_{x\to+\infty} \phi(x)\exp(-Ax) = 0$:*
a) La fonction

$$h(s) = \mathbb{E}\Big(\phi(A_s - N_{D_s} + N_D - D + D_s - C_s)\Big) \tag{35}$$

est décroissante.
b) Par conséquent

$$\mathbb{E}\Big(\phi(A_t - C_t)\Big) \leq \mathbb{E}\Big(\phi(N_D - D)\Big). \tag{36}$$

L'hypothèse portant sur la dérivée seconde de ϕ est contraignante. En effet, on ne peut pas avoir (36) avec $\phi(x) = e^{-\lambda x}$, ainsi le théorème 13 ne permet pas d'obtenir des inégalités de concentration à gauche. Le théorème suivant résout partiellement ce problème, il compare au sens des fonctions convexes $C_t - A_t$ avec une variable aléatoire gaussienne.

Théorème 14 *Soit $(A_s)_{s\geq 0}$ un processus de sauts purs. Supposons qu'il existe une constante $D > 0$ tel que $\|D_t\|_\infty \leq D$. Alors pour toute fonction convexe ϕ de classe C^2 dont la dérivée seconde est croissante et tel qu'il existe un $A > 0$ pour lequel $\lim_{x\to+\infty} \phi(x)\exp(-Ax) = 0$:*
a) La fonction

$$h(s) = \mathbb{E}\Big(\phi(C_s - A_s + W_D - W_{D_s})\Big) \tag{37}$$

est décroissante.
b) Par conséquent,

$$\mathbb{E}\Big(\phi(C_t - A_t)\Big) \leq \mathbb{E}\Big(\phi(W_D)\Big). \tag{38}$$

Comme signalé ci-dessus, les inégalités de concentration convexe permettent d'obtenir des inégalités de concentration classiques. Liptser et Shiryaev [32], Courbot [20] et [21] ont obtenu des inégalités de concentration exponentielles pour des processus à temps continu en bornant la transformée de Laplace. Citons aussi les travaux de Wu [52], Houdré et Privault [25] et Reynaud-Bourret [44], [45] qui obtiennent aussi par des méthodes log-Sobolev des inégalités de concentration exponentielles pour des fonctionnelles de processus de Poisson.

Troisième partie

Outre les maxima de processus cités précédemment, il m'a semblé naturel de chercher des exemples d'applications du théorème 10 du côté des structures combinatoires. Certaines d'entres elles mettent en jeu des processus de comptage $(Z_n)_{n\in\mathbb{N}}$ vérifiant (30) : la plus grande sous-suite croissante dans une permutation aléatoire, le nombre de cliques dans un graphe aléatoire, la hauteur d'un arbre binaire de recherche. L'application du théorème 10 apporterait alors des bornes non asymptotiques. Pour les petites valeurs de n, l'hypothèse 1 peut se vérifier à la main. Il ne semble pas immédiat d'établir ce résultat pour tout entier n. C'est pourquoi il m'a paru néccessaire d'approndir l'analyse probabiliste de l'un de ces exemples.

Ce travail, en collaboration avec Chauvin, Marckert et Rouault, constitue la troisième partie de cette thèse.

Dans cette partie nous étudions les propriétés asymptotiques des arbres binaires de recherche (ABR) dans le modèle dit des permutations aléatoires. On définit

$$\mathbb{U} = \maltese \cup \bigcup_{n\geq 1} \{0,1\}^n$$

l'ensemble des mots finis sur l'alphabet $\{0,1\}$. Si u et v sont dans $\mathbb{U}$, on note uv la concaténation du mot u avec le mot v. On dira alors que uv est un descendant de u et que u est un ancêtre de uv.

Un arbre binaire complet est un sous-ensemble fini de $\mathbb{U}$ vérifiant

$$\begin{cases} \maltese \in T \\ \text{si } uv \in T \text{ alors } u \in T\,, \\ u1 \in T \Leftrightarrow u0 \in T\,. \end{cases}$$

Les éléments de T sont appelés noeuds, et $\maltese$ est nommée la racine. On appelle feuille un élément de T qui n'a pas de descendants. Les noeuds qui ne sont pas des feuilles sont appelés noeuds internes. Un arbre binaire de recherche est un arbre binaire dont les noeuds internes sont étiquetés par des nombres réels, tel que l'étiquette de tout noeud est supérieure aux étiquettes des noeuds de son sous-arbre gauche et inférieure à celles des noeuds de son sous-arbre droit. La profondeur d'un noeud u sera notée dans la suite $|u|$. Ces arbres sont utilisés en informatique pour stocker de façon efficace des données numériques (ou des éléments d'un ensemble muni d'une relation d'ordre total).

Supposons que l'on veuille stocker les réels $a_1, a_2, \ldots$. On procède de la manière suivante : le nombre a_j que l'on insère est comparé à l'étiquette de la racine. S'il n'y a pas encore d'étiquette, on donne à la racine l'étiquette a_j. Si a_j est plus grand (resp. plus petit) que l'étiquette de la racine, on recommence cette procédure dans le sous-arbre droit (resp. gauche). Au temps 0, l'arbre est réduit à ✠ sans étiquette. A l'instant n, on a donc construit un arbre binaire comportant n noeuds internes étiquetés et $n+1$ feuilles. Chacune de ces feuilles est succeptible d'accueillir la $n+1^{\text{ème}}$ donnée (voir la Figure 1 pour la construction d'un ABR jusqu'à $n=5$).

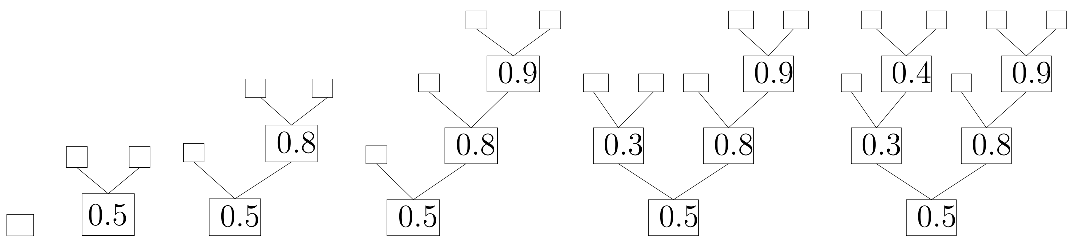

Fig. 1 – ABR construit avec les données 0.5, 0.8, 0.9, 0.3, 0.4 (les carrés vides sont les feuilles).

Dans le modèle dit des permutations aléatoires on suppose que les données $(a_i)_{i\in\mathbb{N}^\star}$ sont des variables aléatoires indépendantes identiquement distribuées, dont la loi est à densité (pour éviter les collisions). Il est évident que seuls les ordres relatifs des $(a_i)_{i\in\mathbb{N}^\star}$ interviennent dans la construction de l'arbre (et non pas les valeurs des a_i). C'est pourquoi, on prendra dans la suite la loi uniforme sur $[0,1]$ et que l'on oubliera les valeurs des étiquettes pour s'intérresser uniquement à la structure arborescente de l'ABR.

Soit σ l'élément de $\mathcal{S}_n$ (ensemble des permutations de $\{1,\dots,n\}$) tel que $a_{\sigma(1)} < a_{\sigma(2)} < \dots < a_{\sigma(n)}$. Par symétrie du problème, on voit que, si les $(a_i)_{i\in\mathbb{N}^\star}$ sont indépendantes et identiquement distribuées (i.i.d.) alors σ est uniforme sur $\mathcal{S}_n$ (ceci justifie la terminologie "modèle des permutations aléatoires"). Toujours pour des raisons de symétrie l'insertion de a_{n+1} se fait de façon uniforme sur les $(n+1)$ feuilles. Pour plus de détails sur les ABR, nous renvoyons le lecteur à l'ouvrage de Mahmoud [37].

On note $U_k(n)$ le nombre de feuilles de l'arbre qui à l'instant n sont à la profondeur k (dans l'ABR représenté par la Figure 1 on a $U_0(5)=0$, $U_1(5)=0$, $U_2(5)=2$ et $U_3(5)=4$). On appelle "profil" de l'arbre à l'instant n la suite des $(U_k(n))_k$.

Ce profil peut être codé par le polynôme aléatoire $\sum_{k\geq 0} U_k(n)z^k$ (de degré égal à la hauteur de l'arbre) ; et d'espérance $C_n(z) = \prod_{k=0}^{n-1}(k+2z)/(k+1)$.

Jabbour [26] a montré que la suite définie par

$$\mathcal{M}_n(z) = C_n(z)^{-1}\sum_{k\geq 0} U_k(n)z^k \tag{39}$$

est une martingale (positive pour $z > 0$). Dans cette troisième partie, on s'intéresse en particulier aux asymptotiques de la martingale $\mathcal{M}_n(z)$ (selon les valeurs de z), ce qui fournit un outil efficace pour l'étude du comportement asymptotique de l'ABR.

Jabbour et ses coauteurs ([19], [26]) ont montré que si $z>0$, la limite p.s. $\mathcal{M}_\infty(z)$ de la martingale positive $\mathcal{M}_n(z)$ est p.s. strictement positive si $z \in (z_c^-, z_c^+)$, avec

$$z_c^- = c'/2 = 0.186..., \quad z_c^+ = c/2 = 2.155... \tag{40}$$

et elle est nulle pour $z \notin [z_c^-, z_c^+]$. Les valeurs critiques z_c^- et z_c^+ sont solutions de l'équation

$$z \log(z/2) - z + 1 = 0.$$

Il utilise ensuite cette martingale pour montrer qu'asymptotiquement le profil de l'arbre a la forme de la courbe de Gauss. Il prouve également que

$$\lim_n \frac{U_k(n)}{\mathbb{E}\big(U_k(n)\big)} = \mathcal{M}_\infty(z), \tag{41}$$

pour $k = 2z \log n + o(\log n)$ et $z \in [0.6, 1.4]$. La méthode utilisée par Jabbour et al. repose sur des techniques de calculs L^2. Cette méthode ne permet pas d'obtenir des informations sur la limite de la martingale sur toute la plage $]z_c^-, z_c^+[$ ni aux points critiques z_c^- et z_c^+ car la martingale ne converge pas L^2 sur toute la plage.

Dans la troisième partie de cette thèse, nous poursuivons et complètons l'étude de cette famille de martingales.

- Nous montrons que $\mathcal{M}_\infty(z) = 0$ p.s. pour $z = z_c^-$ et $z = z_c^+$.
- Nous précisons le comportement aux points critiques en étudiant la convergence de la martingale dérivée

$$\mathcal{M}'_n(z) = \frac{d}{dz}\mathcal{M}_n(z),$$

qui n'est plus positive p.s..

- Nous étendons le domaine des z pour lequel (41) est valable à tout l'intervalle (z_c^-, z_c^+).

Notre approche est différente de celle utilisée par Jabbour et al.. Constatant que certains calculs et résultats ressemblaient à ceux obtenus pour des marches aléatoires de branchement, nous avons mis en évidence une relation structurelle entre l'ABR et un arbre de branchement. Dans les paragraphes suivants, nous allons décrire succinctement les deux points clés de notre approche : le **plongement** et le **tiltage**.

Quelques mots sur le plongement

L'idée de plonger des modèles à temps discret (comme des modèles d'urnes) dans des processus de branchements à temps continu remonte à Athreya-Karlin [4] (voir aussi les travaux de Athreya et Ney [5], Janson [27]). Pour l'ABR de nombreux plongement existent, on pourra regarder par exemple les travaux de Pittel [43], de Biggins-Grey [12], de Devroye [23] ou de Arratia et al. [2].

Nous plongeons l'ABR dans un processus à temps continu de manière à créer de l'indépendance entre sous-arbres disjoints. Le remplacement à l'étape $n+1$ d'une feuille présente à l'instant n par un noeud interne et deux feuilles suggère naturellement une fission binaire.

Nous introduisons un processus $(\mathbb{T}_t)_{t \geq 0}$ à valeurs arbres binaires que nous codons au moyen de la fragmentation de $]0,1[$ suivante (voir Aldous-Shields [1]).

<u>Processus de fragmentation de $]0,1[$</u> :

Soit F l'ensemble des ouverts de $]0,1[$. On construit un processus $(F(t))_{t\geq 0}$ à temps continu que l'on nomme processus de fragmentation de $]0,1[$.

Au temps $t=0$, on pose $F(0)=]0,1[$. Au temps t, $F(t)$ est composé d'un certain nombre de sous-intervalles ouverts et disjoints de $]0,1[$, chacun muni d'une horloge exponentielle de paramètre 1 (indépendante des autres).

Lorsqu'une horloge sonne, l'intervalle correspondant se scinde en deux, et on munit à nouveau chacun de ces intervalles d'une horloge neuve.

Ainsi $F(0)=(0,1)$, $F(\tau_1)=\{(0,1/2),(1/2,1)\}$ où τ_1 est une variable aléatoire de loi exponentielle de paramètre 1, ainsi de suite. On note $\tau_1,\tau_2,\ldots$ les instants de fragmentations.

On peut interpréter les deux fragments issus d'un même fragment I comme étant les fils gauche et droit de I. On voit apparaître ainsi une structure arborescente, et même un processus à valeurs arbres (voir la Figure 2).

L'arbre généalogique $(\mathbb{T}_t, t\geq 0)$ associé à cette fragmentation (chaque fragment étant codé par un noeud de l'arbre) est appelé arbre de Yule. On notera $\partial\mathbb{T}_t$ les feuilles de $\mathbb{T}_t$ à l'instant t.

On obtient alors simplement que l'arbre de Yule arrêté aux temps $(\tau_i, i\geq 0)$ à la même loi que l'ABR. De plus ces deux arbres peuvent être construit sur le même espace (voir aussi Kingman [28]).

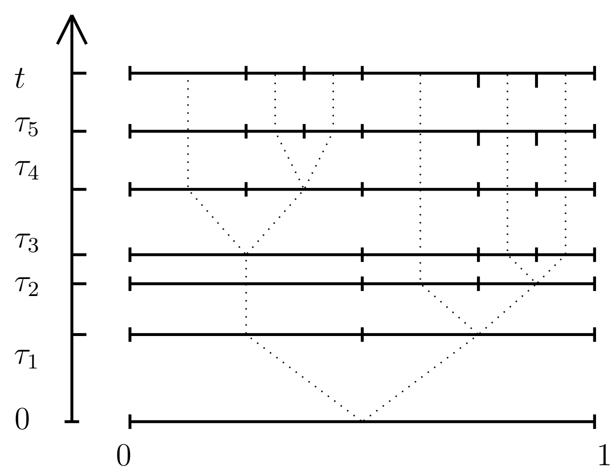

Fig. 2 – Fragmentation et arbre de Yule.

La hauteur $|v|$ d'une feuille v de $\mathbb{T}_t$ est égale à $-\log_2 I_v$ où I_v est l'intervalle ouvert de $F(t)$ codé par v.

Le processus à valeurs mesures $\sum_{v\in\partial\mathbb{T}_t}\delta_{|v|}$ apparait alors comme une marche aléatoire de branchement à temps continu (Uchiyama [51], Biggins [11] Kyprianou [29]). Après reparamétrage , la famille associée de martingales additives peut s'écrire sous la forme

$$M(t,z) := \sum_{u\in\partial\mathbb{T}_t} z^{|u|}e^{t(1-2z)},$$

ce qui donne la relation

$$M(\tau_n,z) = e^{\tau_n(1-2z)}C_n(z)\mathcal{M}_n(z)\,.$$

Grâce à cette formule, on peut transférer certains des résultats connus sur $(M(t,z), t \geq 0)$ à $(\mathcal{M}_n(z), n \geq 0)$. On obtient alors une nouvelle preuve de certains résultats de Jabbour sur la limite de la martingale $\mathcal{M}_n(z)$. De plus, cette relation entre les deux martingales nous permet de déterminer la limite de la martingale $\mathcal{M}_n(z)$ aux deux points critiques z_c^- et z_c^+, et d'étudier la martingale dérivée.

Quelques mots sur le tiltage

Cette méthode consiste à interpréter $(M(t,z), z \geq 0)$ et $(\mathcal{M}_n(z), n \geq 0)$ comme des dérivées de Radon-Nikodym, et à décrire précisément le comportement des arbres et des fragmentations sous la nouvelle probabilité.

La méthode du **tiltage** consiste à marquer aléatoirement une des branches de l'arbre (plus exactement une branche "infinie"). On procède de la manière suivante :

On se place dans le modèle de fragmentation. On se donne une variable aléatoire U uniforme sur $]0,1[$, indépendante de la fragmentation et on note $V(t)$ l'intervalle de $F(t)$ qui contient U. On dit alors que l'intervalle $V(t)$ est marqué.

Comme ci-dessus, tous les intervalles de $F(t)$ sont munis d'une horloge exponentielle de paramètre 1 sauf l'intervalle $V(t)$ que l'on muni d'une horloge exponentielle de paramètre $2z$.

Ainsi l'intervalle $V(t)$ ne se scinde pas à la même vitesse que les autres intervalles (il se scinde plus rapidement si $2z > 1$ et moins rapidement sinon).

Comme on l'a dit ci-dessus on peut interpréter le processus de fragmentation comme un processus à valeurs arbres. Les noeuds de l'arbre correspondant à l'intervalle marqué constituent une branche que l'on appelle "épine d'orsale" (**spine**) (dans la Figure 3, le spine correspond aux noeuds rouges). Dans la suite on appellera cette branche marquée *spine*. L'évolution de l'ABR peut se décomposer en l'évolution du spine et en l'évolution des sous-arbres qui lui sont accrochés (ce sont les classes de la Figure 3). Ces classes peuvent s'interpréter comme les tables d'un restaurant chinois de Dubins et Pitman [42]. Ainsi des propriétés de l'arbre se trouvent réduites à des propriétés le long d'une seule branche, du coup de type "somme de variables aléatoires indépendantes".

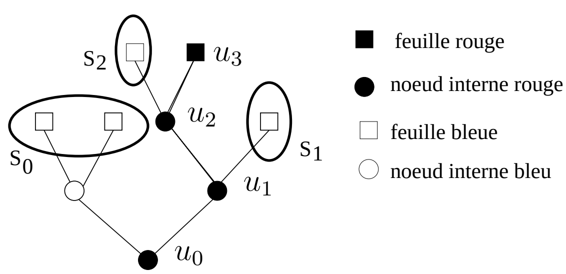

FIG. 3 – Un arbre tilté et ces différentes classes.

Ceci, nous permet dans le Chapitre 5, de présenter une preuve "conceptuelle" de la convergence de $\mathcal{M}_n(z)$, dans la terminologie de Lyons, Pemantle et Peres[33, 34] et de démontrer la convergence du profil normalisé

$$\frac{U_k(n)}{\mathbb{E}(U_k(n))}.$$

Cette méthode (du *tiltage*) a été utilisée pour les marches branchantes (Biggins et Kyprinou [13]), pour les fragmentations (Bertoin et Rouault [10]) ainsi que pour d'autres processus (Athreya [3], Barlow, Pemantle et Perkins [7]).

Bibliographie

[1] D. Aldous, P. Shields. A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields,* 79 : 509-542. 1988.

[2] R. Arratia, A. Barbour, S.Tavaré. Logarithmic Combinatorial Structures : a Probabilistic Approach. *Preprint (book) available at http ://www-hto.usc.edu/books/tavare/index.html.*

[3] K. Athreya. Change of measures for Markov chains and the L log L theorem for branching processes. *Bernoulli.* vol. 6, 323-338 1999.

[4] K. B. Athreya, S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist. 39 : 1801-1817.* 1968.

[5] K. B. Athreya, K. P. E. Ney. Branching processes. *Springer-Verlag, New-York.* 1972.

[6] K. Azuma. Weighted sums of certain dependent random variables *Tohoku Math. J., II. Ser. 19, 357-367.* 1967

[7] M. T. Barlow, R. Pemantle, E. A. Perkins. Diffusion-limited aggregation on a tree. *Probab. Theory Relat. Fields.* 107 : 1-60. 1997.

[8] G. Bennett. Probability inrqualities for the sum of independent random variables . *J. of the American Statistical Association.* 57 : 33-45. 1962.

[9] V. Bentkus. *on Hoeffding's inequality* To appear in Ann. of Probab. 2003

[10] J. Bertoin, A. Rouault. Additive martingales and probability tilting for homogeneous fragmentations. *Preprint PMA-808 available at http ://www.proba.jussieu.fr/mathdoc/preprints/index.html.* 2003.

[11] J. D. Biggins. Uniform convergence of martingales in the branching random walk. *Ann. Probab.* 20, 1 :137-151. 1992.

[12] J. D. Biggins, D. R. Grey. A note on the growth of random tress. *Statist. Probab. Lett.* 32, 4 : 339-342. 1997.

[13] J. D. Biggins, A. E. Kyprianou. Measure change in multitype branching. *Preprint available at http ://www.shef.ac.uk/ st1jdb/mcimb.html.* 2001.

[14] L. Birgé, P. Massart. Minimum contrast estimators on sieves : Exponential bounds and rates of convergence. *Bernoulli 4, No.3, 329-375* 1998.

[15] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris, Ser I 334* 495-500. 2002.

[16] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. *To appear in the proceedings of the conference on stochastic inequalities. Eds Giné, Houdré and Nualart.* 2002.

[17] P. Brémaud. Point processes and queues. Springer-Verlag. 1981.

[18] J. Bretagnolle. Statistique de Kolmogorov-Smirnov pour un échantillon non équiréparti. Dans aspects statistiques et aspects physiques des processus gaussiens. *Paris : Editions du centre nationnal de la recherche scientifique.* 1981.

[19] B. Chauvin, M. Drmota, J. Jabbour-Hattab. The profile of Binary Search Trees. *Ann. Appl. Prob.* 11 : 1042-1062. 2001.

[20] B. Courbot. Vitesses de convergence pour les martingales dans le théorème central limite fonctionnel : comparaisons de méthodes, cadres uni et multidimentionnel. Thèse de doctorat de l'université de Rennes 1, 1998

[21] B. Courbot. Rates of convergence in the functional CLT for martingales. *C.R. Acad. Sci. Paris, t. 328, Série 1, p. 509-513,* 1999.

[22] C. Dellacherie, P.A. Meyer. Probabilités et potentiel. Chap. 5 to 8. Hermann, Paris. 1980.

[23] L. Devroye. Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics.* M. Habib et al. Springer. 1998.

[24] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc* 58, 13-30. 1963.

[25] C. Houdré, N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli 8, No.6, 697-720.* 2002

[26] J. Jabbour-Hattab. Martingales and Large Deviations for Binary Search Trees. *Random Structure and Algorithms* 19 : 112-127. 2001.

[27] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Preprint available at http ://www.math.uu.se/∼svante/papers/index.html.*

[28] J. F. C. Kingman. The coalescent process. *Stochastic Process. Appl. 13 : 235-248.* 1982.

[29] A. E. Kyprianou. A note on branching Lévy processes *Stochastic Process. Appl.* 82, 1, 1-14. 1999.

[30] M. Ledoux. On Talagrand's deviation inequalities for product measures. ESAIM *probability and statistics*, 1 :63-87. 1996.

[31] M. Ledoux. The Concentration of measure Phenomenon *Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI.* 2001.

[32] R.S. Liptser, A.N. Shiryaev. Theory of martingales. Kluwer Academic Publishers, Dordrecht, 1986.

[33] R. Lyons. A simple path to Biggins' martingale convergence for the branching random walk. *Classical and Modern Branching Processes. K.B. Athreya, P. Jagers. IMA Volumes in Mathematics and its Applications. Springer. 84. 217-222.* 1997.

[34] R. Lyons, R. Pemantle, Y. Peres. Conceptual proofs of L log L criteria for mean behavior of branching processes. *Ann. Probab. 23 : 1125-1138.* 1995.

[35] C. McDiarmid. On the method of bounded differences. *Surveys in combinatorics, Inv. Pap. 12th Br. Comb. Conf., Norwich/UK 1989, Lond. Math. Soc. Lect. Note Ser. 141, 148-188.* 1989.

[36] C. McDiarmid. Concentration., *Habib, Michael (ed.) et al., Probabilistic methods for algorithmic discrete mathematics. Berlin : Springer. Algorithms Comb. 16, 195-248.* 1998.

[37] H. Mahmoud. Evolution of Random Search Trees. *John Wiley.* 1992.

[38] P. Massart. About the constant in Talagrand's concentration inequalities for empirical processes. *Annals of Probab.* 28 :863-884. 2000.

[39] P. Massart. Cours de Saint-Flour 2002 : Concentration inequalities and Model Selection. *To be published.* 2003.

[40] I. Pinelis. Optimal Bounds for the Distributions of Martingales in Banach Spaces. *Ann. Prob.* Vol. 22. No. 4, 1679-1706 1994

[41] I. Pinelis. Optimal Tail Comparison Based on Comparison Moments *Eberlein, Ernst (ed.) et al., High dimensional probability. Proceedings of the conference, Oberwolfach, Germany, August 1996. Basel : Birkhäuser. Prog. Probab.* Vol. 43, 297-314 1998.

[42] J. Pitman. Cours de Saint-Flour 2002 : Combinatorial Stochastic Processes. *Available at http ://stat-www.berkeley.edu/users/pitman/bibliog.html.* 2002.

[43] B. Pittel. On growing random binary trees. *J. Math. Anal. Appl.* 103, 2 : 461-480. 1984.

[44] P. Reynaud-Bourret. Exponential inequalities for counting processes. *preprint of the Georgia Institute of Technology.* 2002.

[45] P. Reynaud-Bourret. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theor. relat. Fields.* 126, 103-153 2003.

[46] E. Rio. Théorie asymptotique des processus aléatoires faiblement dépendants. J.M. Ghidaglia, X. Guyon (Ed.). Mathématiques et Applications 31. Springer. 2000.

[47] E. Rio. Une inégalité de Bernstein pour les maxima de processus empiriques. *Preprint 57 de l'Université de Versailles–Saint-Quentin.* 2001.

[48] Q.M. Shao. A Comparison Theorem on Moment Inequalies Between Negatively Associated and Independent Random Variables. *Journ. of Theor. Prob.*, 2000. Vol. 13, 2. 343-356.

[49] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math. 126 :503-563,* 1996.

[50] S. Tavaré. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.* 25, 2, 161-168. 1987.

[51] K. Uchiyama. Spatial growth of a branching process of particles living in R. *Ann. Probab.* 10, 4 : 896-918 1982.

[52] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Relat. Fields 119, 427-438*, 2000.

Deuxième partie

Inégalités de Concentration de type Talagrand

Chapitre 1

Inégalités de concentration à gauche de type Talagrand dans le cas i.i.d

Une inégalité de concentration à gauche pour les processus empiriques

Thierry KLEIN
Université de Versailles-Saint-Quentin
45 avenue des Etats-Unis
78035 Versailles cedex
klein@math.uvsq.fr

Résumé 1 *Nous donnons des constantes dans l'inégalité de concentration à gauche de Talagrand pour les processus empiriques indexés par des classes de fonctions, en partant de la méthode de Herbst. Le point nouveau est que la constante du facteur variance est exacte, ce qui répond à une conjecture de Massart.*

Abstract 1 *We give new constants in Talagrand's left concentration inequality for maxima of empirical processes. Our approach is based on the Herbst method. The improvement we get concerns the constant in the variance factor, which is the one conjectured by Massart.*

1.1 Introduction

Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes de loi commune P à valeur dans $\mathcal{X}$ espace polonais muni de sa tribu borélienne. Soit $\mathcal{F}$ une famille dénombrable de fonctions mesurables de $\mathcal{X}$ dans $\mathbb{R}$, de carré intégrable sous P. On pose

$$S_n(f) = f(X_1) + \ldots + f(X_n). \tag{1.1}$$

Nous regardons les propriétés de concentration de la variable aléatoire $Z = \sup\{S_n(f) : f \in \mathcal{F}\}$ autour de sa moyenne. Ce problème a été étudié dans une série de travaux successifs, en particulier par Talagrand [6] au moyen d'inégalités isopérimétriques pour les mesures produits. Il a obtenu une inégalité de type Bennett pour la déviation de Z par rapport à sa moyenne. Ensuite Ledoux [2] a montré que ces inégalités pouvaient aussi être obtenues à partir d'inégalités de type log-Sobolev pour les mesures produit. Cette technique permet de majorer la transformée de Laplace de Z. Massart [3] a clairement montré l'intérêt des méthodes entropiques (ou log-Sobolev) pour le calcul des constantes dans les inégalités de Talagrand, enfin Rio [5] a obtenu une inégalité de concentration à droite de type Bernstein avec un facteur variance asymptotiquement exact. Nous allons montrer dans cette note un analogue de l'inégalité de Rio [5] pour la concentration de Z à gauche de sa moyenne. Bousquet [1] a amélioré la fonction de taux, et a obtenu une inégalité de Bennett.

1.2 Résultat

Théorème 1 *Soit $\mathcal{F}$ classe dénombrable de fonctions de $\mathcal{X}$ dans $]-\infty, 1]$, mesurables, de carré intégrable et d'espérance nulle sous P. Soit σ^2 telle que $\sigma^2 \geq P(f^2)$ pour toute f dans $\mathcal{F}$.*

- *Si les fonctions de $\mathcal{F}$ sont à valeurs dans $[-1, 1]$ alors pour tout $x > 0$,*

$$P\big(Z \leq E(Z) - x\big) \leq \exp\left(- \frac{v_n}{16} h(\frac{4x}{v_n})\right), \tag{1.2}$$

avec $v_n = n\sigma^2 + 2E(Z)$ et $h(x) = (1 + x)\log(1 + x) - x$.

- *Si pour tout $f \in \mathcal{F}$, et tout $p \geq 2$, $\mid E(f^p(X_i)) \mid \leq \frac{\sigma^2 p!}{2}$ alors*

$$\forall t \in [0, 1[, \qquad L(t) \leq -te^{-t}E(Z) + \frac{n\sigma^2}{2}\frac{t^2(1 + 2t)}{1 - t}, \tag{1.3}$$

et par conséquent pour tout x positif,

$$P(Z \leq E(Z) - x) \leq \exp\Big(- (u - v)^2\big(1 - 2\frac{v(u - v)}{u^2}\big)\Big), \tag{1.4}$$

où $L(t)$ est la log-*Laplace de $-Z$, $u = \sqrt{x + v_n/2}$, $v = \sqrt{v_n/2}$.*

Remarque 1 *La fonction de taux de $v_n t^2(2(1-t))^{-1}$ est $(u - v)^2$, et nous avons en plus une correction multiplicative comprise entre $1/2$ et 1 tendant vers 1 quand xv_n^{-1} tend vers 0. Asymptotiquement, on retrouve donc la fonction de taux des inégalités gaussiennes.*

1.3 Preuve

1.3.1 Lemme préparatoire

Nous rappelons ici le Lemme 2.1 de Rio [5] (voir aussi Massart [3], lemme 8).

Lemme 1 *Soient $(X_1, \ldots, X_n)$ des variables aléatoires indépendantes à valeurs dans un espace polonais $\mathcal{X}$. Soit $\mathcal{F}_n^k$ la tribu engendrée par $(X_i)_{i\neq k}$. Posons $\phi(x) = x\log(x) - x + 1$. Soit h une variable aléatoire fonction mesurable de $(X_1, \ldots, X_n)$, strictement positive et intégrable. Alors, pour toute famille (h_k) de variables aléatoires réelles, strictement positives et intégrables, respectivement $\mathcal{F}_n^k$-mesurables,*

$$E(h\log h) - E(h)\log E(h) \leq \sum_{k=1}^{n} E(h_k\phi(h/h_k)). \tag{1.5}$$

1.3.2 Preuve

Comme dans Rio [5], on peut supposer que $\mathcal{F}$ est finie et procéder ensuite par passage à la limite. On définit $\mathcal{F}_n^k$ la tribu engendrée par $(X_i)_{i\neq k}$, $Z_k = \sup\{S_n^k(f) : f \in \mathcal{F}\}$. Puisque $\mathcal{F}$ est finie, on peut opérer une sélection $\mathcal{F}_n^k$−mesurable d'une fonction f_k telle que $Z_k = S_n^k(f_k)$, $Z'_k = S_n(f_k)$. On définit $\psi(x) = 1-(1+x)e^{-x}$, $\eta_k = f_k(X_k)$. Soit $t > 0$, en appliquant (1.5) à $h = e^{-tZ}$ et $h_k = e^{-tZ_k}$, on obtient l'inégalité suivante pour la transformée de Laplace F de $-Z$.

$$tF'(t) - F(t)\log F(t) \leq \sum_{k=1}^n E\big(e^{-tZ_k}\psi(t(Z-Z_k))\big). \tag{1.6}$$

On note alors que

$$\psi(tx) = \frac{t^2}{1-t}xe^{-tx} + \Big(1 - e^{-tx}(\frac{1-t+tx}{1-t})\Big) := q(t)xe^{-tx} + r(t,x), \tag{1.7}$$

si $q(t) = t^2/(1-t)$. Par conséquent (1.6) s'écrit encore

$$tF'(t) - F(t)\log F(t) \leq \sum_{k=1}^n E\big(e^{-tZ}\frac{t^2}{1-t}(Z-Z_k)\big) + \sum_{k=1}^n E\big(e^{-tZ_k}r(t,Z-Z_k)\big).$$

Comme $\sum_{k=1}^n (Z-Z_k) \leq Z$ et $q(t) > 0$, on en déduit que

$$tF'(t) - F(t)\log F(t) \leq -q(t)F'(t) + \sum_{k=1}^n E\Bigg(e^{-tZ_k}r(t,Z-Z_k)\Bigg). \tag{1.8}$$

Puisque

$$\frac{\partial r}{\partial x}(t,x) = \frac{t^2}{1-t}e^{-tx}(x-1),$$

la fonction $r(t,x)$ est décroissante en x pour $0 \leq t < 1$ et $x \leq 1$.

Puisque $\eta_k = f_k(X_k) \leq (Z-Z_k) \leq 1$ et r est décroissante en x,

$$tF' - F\log F \leq -q(t)F' + \frac{1}{1-t}\sum_{k=1}^n E\Bigg(e^{-tZ_k}\Big(1-t-e^{-t\eta_k}(1-t+t\eta_k)\Big)\Bigg). \tag{1.9}$$

Posons $S_k = 1-t-e^{-t\eta_k}(1-t+t\eta_k)$. L'équation (1.9) devient alors pour $0 \leq t < 1$,

$$tF'(t) - (1-t)F(t)\log F(t) \leq \sum_{k=1}^n E\Big(e^{-tZ_k}E_n^k(S_k)\Big), \tag{1.10}$$

où E_n^k est l'espérance conditionnelle par rapport à $\mathcal{F}_n^k$.

• ***Sous la première hypothèse*** : montrons à l'aide d'un développement en série entière de S_k que :

$$S_k \leq -t^2\eta_k + \eta_k^2(e^t - 1 - t + t^3). \tag{1.11}$$

Preuve de (1.11).

On pose pour la preuve $x = \eta_k$. On a alors, pour $t \in [0,1]$,

$$\begin{aligned} S_k =&(1-t) - (1-t+tx)(1-tx+t^2x^2/2) - (1-t)\sum_{j=3}^{+\infty}(-1)^j x^j t^j/j! \\ &\qquad - tx\sum_{j=3}^{+\infty}(-1)^j x^j t^j/j!\ , \\ =& - t^2x + t^2x^2/2 + t^3\frac{x^2}{2}(1-x) + (1-t)x^2\sum_{j=3}^{+\infty}(-1)^{j+1}x^{j-2}t^j/j! \\ &\qquad + tx^2\sum_{j=3}^{+\infty}(-1)^{j+1}x^{j-1}t^j/j!\ . \end{aligned}$$

Comme x est dans $[-1,1]$, on a la majoration suivante :

$$\begin{aligned} S_k &\le -t^2x + t^2x^2/2 + t^3x^2 + (1-t)x^2(\frac{e^t - e^{-t}}{2} - t) + tx^2(e^t - 1 - t - \frac{t^2}{2}), \\ &\le -t^2x + x^2\Big(\frac{t^2}{2} + \frac{t^3}{2} - 2t + e^t(\frac{1+t}{2}) - e^{-t}(\frac{1-t}{2})\Big). \end{aligned}$$

Notons maintenant

$$\begin{aligned} \alpha(t,x) &= -t^2x + x^2\Big(\frac{t^2}{2} + \frac{t^3}{2} - 2t + e^t(\frac{1+t}{2}) - e^{-t}(\frac{1-t}{2})\Big), \\ \beta(t,x) &= -t^2x + x^2\Big(e^t - 1 - t + t^3\Big), \\ \gamma(t,x) &= \beta(t,x) - \alpha(t,x), \end{aligned}$$

et montrons que $\gamma \ge 0$. On remarque que le signe de γ est le même que celui de

$$\delta(t) = e^t(1-t) + e^{-t}(1-t) - 2 + 2t + t^3 - t^2.$$

comme $\delta(0) = 0$, $\delta'(t)e^t = -te^{-2t} + (2 + 3t^2 - 2t)e^t + (t-2)$, $\delta'(0) = 0$ et $\delta'(1) > 0$ et comme la fonction de $q(y) = -ty^2 + (2 - 2t + 3t^2)y + (t-2)$ est concave on en déduit que δ est positive, ce qui achève la preuve de (1.11).

Par (2.4) dans Rio [5] on sait que $E_n^k(\eta_k) = 0$ et $E_n^k(\eta_k^2) \le \sigma^2$. On en déduit la majoration suivante :

$$E_n^k(S_k) \le \sigma^2(e^t - 1 - t + t^3). \tag{1.12}$$

En remarquant que $e^{-tZ_k} \le e^t E\big(e^{-tZ}\big)$, on obtient l'inéquation différentielle suivante valable pour tout $0 \le t < 1$:

$$tL'(t) - (1-t)L(t) \le ne^t\sigma^2(e^t - 1 - t + t^3). \tag{1.13}$$

Montrons maintenant que la fonction

$$G(t) = n\sigma^2 \frac{e^{4t} - 1 - 4t}{16} - te^{-t}E(Z) \tag{1.14}$$

est une sursolution de (1.13) sur $[0, 1]$. Soit

$$\begin{aligned} f(t) &= tG'(t) - (1-t)G(t) \\ &= \frac{n\sigma^2}{16}\Big(4t(e^{4t} - 1) - (1-t)(e^{4t} - 1 - 4t)\Big). \end{aligned}$$

Il s'agit donc de montrer que, pour tout t dans $[0, 1[$,

$$f(t) \geq n\sigma^2 e^t(e^t - 1 - t + t^3). \tag{1.15}$$

Cela revient à montrer que , pour tout t dans $[0, 1[$

$$D(t) = e^{4t}(5t - 1) + 1 - t - 4t^2 - 16e^{2t} + 16e^t(1 + t - t^3) \geq 0.$$

Nous avons les formules suivantes pour les dérivées de D,

$$\begin{aligned} D' &= e^{4t}(2t + 1) - 1 - 8t - 32e^{2t} + 16e^t(2 + t - 3t^2 - t^3), \\ D^{(2)} &= 8e^{4t}(10t + 3) - 8 - 64e^{2t} + 16e^t(3 - 5t - 6t^2 - t^3), \\ D^{(3)} &= 16e^{4t}(5t + 11) - 128e^{2t} - 16e^t(2 + 17t - 9t^2 - t^3). \end{aligned}$$

Montrons que $D^{(3)}$ est positive. Soit

$$\begin{aligned} R(t) &= \frac{D^{(3)}(t)}{8e^t}, \\ R(t) &= e^{3t}(11 + 5t) - 8e^t - (2 + 17t - 9t^2 - t^3), \\ R'(t) &= e^{3t}(38 + 15t) - 8e^t - (17 - 18t - 3t^2) > 0. \end{aligned}$$

Donc R est croissante sur $[0, 1[$. Or $R(0) = 1$ donc $R(t) > 0$ sur $[0, 1[$ et $D^{(3)} \geq 0$ sur $[0, 1[$. On en déduit alors facilement que $D(t) > 0$. Ce qui entraîne que $G(t)$ est une sursolution de (1.13).
Comme de plus $G(0) = L(0)$ et $G'(0) = L'(0)$, on en déduit que pour tout $t \in [0, 1[$ on a

$$L(t) \leq G(t). \tag{1.16}$$

Comme sur $[0, 1[$ on a $t(1 - e^{-t}) \leq 2\frac{e^{4t}-1-4t}{16}$, on en déduit que pour $t \in [0, 1[$:

$$P\big(Z \leq E(Z) - x\big) \leq \exp\Big(v_n \frac{e^{4t} - 1 - 4t}{16} - tx\Big). \tag{1.17}$$

Preuve de (1.2) pour $x < v_n(e^4 - 1)/4$.

L'expression $\exp\Big(v_n \frac{e^{4t}-1-4t}{16} - tx\Big)$ est optimisée pour t_0 vérifiant

$$4t_0 = \ln(1 + \frac{4x}{v_n}).$$

En réinjectant t_0 dans (1.17) on obtient (1.2).

Preuve de (1.2) pour $x \geq v_n(e^4 - 1)/4$.
Pour obtenir (1.2) quand $x \geq v_n(e^4 - 1)/4$ il suffit de remarquer que pour tout $f \in \mathcal{F}$ on a

$$P(Z \leq E(Z) - x) \leq P(S_n(f) \leq E(Z) - x), \tag{1.18}$$

et que le membre de droite de (1.18) peut être majoré d'après l'inégalité de Bennett classique, par

$$P(S_n(f) \leq E(Z) - x) \leq \exp\left(- n\sigma^2 h\Big(\frac{x - E(Z)}{n\sigma^2}\Big)\right). \tag{1.19}$$

On conclut en remarquant que pour $x \geq v_n(e^4 - 1)/4$, puisque $v_n/2 \geq E(Z)$,

$$x - E(Z) \geq x(1 - \frac{2}{e^4 - 1}) \geq \frac{24x}{25}.$$

Donc, comme h est croissante, (1.19) devient

$$P(S_n(f) \leq E(Z) - x) \leq \exp\left(- n\sigma^2 h\Big(\frac{24x}{25n\sigma^2}\Big)\right). \tag{1.20}$$

Or le majorant de (1.20)est à nouveau inférieur à la borne de (1.2).

•***Sous la deuxième hypothèse*** : comme $\eta_k = f_k(X_k)$ avec f_k $\mathcal{F}_n^k-$mesurable on a $| E_n^k(\eta_k^p) | \leq \sigma^2 p!/2$. Montrons alors, à l'aide d'un développement en série entière de S_k que :

$$E_n^k\big(S_k\big) \leq \sigma^2 \frac{t^2}{2(1-t)^2}(1 + 2t - 3t^2 + t^3). \tag{1.21}$$

Preuve de (1.21).

Posons $x = \eta_k$. Alors

$$S_k = -t^2 x + t^3 \frac{x^2}{2}(1 - x) + (1 - t)\sum_{i=3}^{+\infty}(-1)^{i+1} x^i \frac{t^i}{i!} + t\sum_{i=3}^{+\infty}(-1)^{i+1} x^{i+1}\frac{t^i}{i!},$$

et par conséquent,

$$\begin{aligned} E_n^k(S_k) &\leq \frac{\sigma^2}{2}\Big(3t^3 + t^2 + (1-t)\sum_{i=3}^{+\infty} t^i + t\sum_{i=3}^{+\infty}(i+1)t^i\Big), \\ &\leq \frac{\sigma^2 t^2}{2(1-t)^2}(1 + 2t - 3t^2 + t^3). \end{aligned}$$

ce qui prouve (1.21).
On obtient alors l'inéquation différentielle suivante, valable pour tout $0 \leq t < 1$:

$$tL'(t) - (1-t)L(t) \leq \frac{n\sigma^2 t^2 e^t}{2(1-t)^2}(1 + 2t - 3t^2 + t^3). \tag{1.22}$$

Cette équation peut s'écrire

$$\left(\frac{L(t)}{te^{-t}}\right)' \leq \frac{n\sigma^2 e^{2t}}{2(1-t)^2}(1 + 2t - 3t^2 + t^3). \tag{1.23}$$

Définissons la fonction G par

$$G(t) = \frac{n\sigma^2 e^t (t + 2t^2)}{2(1-t)}.$$

Alors

$$G'(t) = \frac{n\sigma^2 e^t (1 + 5t - t^2 - 2t^3)}{2(1-t)^2}.$$

Nous allons montrer que pour tout t dans $[0, 1[$,

$$G'(t) \geq \frac{n\sigma^2 e^{2t}}{2(1-t)^2}(1 + 2t - 3t^2 + t^3).$$

Pour cela il suffit de montrer que

$$D(t) = 1 + 5t - t^2 - 2t^3 - e^t(1 + 2t - 3t^2 + t^3) \geq 0. \qquad (\mathcal{H})$$

On a

$$\begin{aligned} D^{(3)}(t) &= -12 - e^t(-5 + 2t + 6t^2 + t^3) \leq 0, \\ D^{(2)}(t) &= -2 - 12t - e^t(-1 - 4t + 3t^2 + t^3),\ D^{(2)}(0) = -1, \\ D'(t) &= 5 - 2t - 6t^2 - e^t(3 - 4t + t^3). \end{aligned}$$

On en déduit que D' est décroissante. $(\mathcal{H})$ est donc une conséquence immédiate de la positivité de la fonction concave D aux bords de l'intervalle. Après majoration et intégration de (1.23) on obtient :

$$L(t) \leq -te^{-t}E(Z) + \frac{n\sigma^2}{2}\frac{t^2(1+2t)}{(1-t)} \leq -tE(Z) + \frac{v_n}{2}\frac{t^2(1+2t)}{1-t}. \tag{1.24}$$

Par le calcul usuel de Cramer-Chernov on obtient

$$P(Z \leq E(Z) - x) \leq \exp\left(\frac{v_n}{2}\frac{t^2(1+2t)}{(1-t)} - tx\right). \tag{1.25}$$

Pour obtenir une borne on prend alors le réel $t^\star$ minimisant $\frac{1}{2}v_n t^2(1-t)^{-1} - tx$. Après calcul on trouve $t^\star = 1 - w_n^{\frac{1}{2}}(x + w_n)^{-\frac{1}{2}}$, où $w_n = v_n/2$ que l'on réinjecte dans (1.25), pour obtenir (1.4) . $\square$

Bibliographie

[1] Olivier Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris, Ser I 334* 2002. 495-500.

[2] Michel Ledoux. On Talagrand's deviation inequalities for product measures. ESAIM *probability and statistics*, 1996. 1 :63-87.

[3] Pascal Massart. About the constant in Talagrand's concentration inequalities for empirical processes. *Annals of Probab.* 2000. 28 :863-884.

[4] Emmanuel Rio. Théorie asymptotique des processus aléatoires faiblement dépendants. J.M. Ghidaglia, X. Guyon (Ed.). Mathématiques et Applications 31. Springer. 2000.

[5] Emmanuel Rio. Une inégalité de Bernstein pour les maxima de processus empiriques. *Preprint 57 de l'Université de Versailles–Saint-Quentin*. 2001.

[6] Michel Talagrand. New concentration inequalities in product spaces. *Invent. Math. 126 :503-563,* 1996.

Chapitre 2

Inégalités de concentration de type Talagrand dans le cas d'indépendance

Empirical processes and concentration around the mean

T. Klein and E. Rio, *UMR 8100 CNRS, Université de Versailles Saint Quentin en Yvelines, 45 Avenue des Etats Unis, 78035 VERSAILLES.*

Abstract 1 In this paper we give optimal constants in Talagrand's concentration inequalities for maxima of empirical processes associated to independent and eventually non identically distributed random variables. These bounds are then applied to concentration inequalities for Rademacher processes. Our approach is based on the entropy method introduced by Ledoux.

M*athematics subject classifications.* (̊1991) - 60 E 15, 60 F 10.

2.1 Introduction

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with values in some Polish space $\mathcal{X}$ and let $\mathcal{S}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]^n$. For s in $\mathcal{S}$, denoting by $s^1, \ldots, s^n$ the components of s, we set

$$S_n(s) = s^1(X_1) + \cdots + s^n(X_n). \tag{2.1}$$

In this paper we are interested in concentration inequalities for $Z = \sup\{S_n(s) : s \in \mathcal{S}\}$.

Now let us recall the main results in this direction. Starting from concentration inequalities for product measures, Talagrand in [8] (1996) obtains Bennett type upper bounds on the Laplace transform of Z. More precisely he proves

$$\log \mathbb{E} \exp(tZ) \leq t\mathbb{E}(Z) + aV_n(e^{bt} - bt - 1) \tag{2.2}$$

for any positive λ. Here V_n is close to the maximal variance of $S_n(s)$. The conjecture concerning the constants is then $a = b = 1$. Neverthless is seems dificult to reach the exact constants via Talagrand's method. In order to obtain concentration inequalities more directly, Ledoux in [3] (1996) uses a Log-Sobolev type method together with a powerfull argument of tensorization of the entropy. When applied to $\exp(tZ)$, this method yields a differential inequality (this is the so-called Herbst argument) on the Laplace transform of Z and gives (2.2) again. This method makes it possible to reach optimal constants, as noted by Massart in [5] (2000), who obtains explicit constants, not far from the optimal constant) in (2.2). Later on, the constants in Talagrand's inequality have been improved in the case of identically distributed random variables by Rio in [7] (2002) who obtains optimal constants in the moderate deviations bandwith, and next by Bousquet in [1] (2002), who obtains (2.2) with $a = b = 1$ for positive values of t. The case of negative values of t has been studied by Klein in [2] (2002), who obtains (2.2) for $b = 4$ and $a = 1/16$ (note that Klein's result is optimal in the moderate deviations bandwith).

Here we are interested in optimal constants in Talagrand's inequalities for non identically distributed random variables. Our approach to obtain the best constants is to apply the lemma of tensorizarion of the entropy method proposed by Ledoux in [3] (1996). The relative entropies are then bounded up via the duality lemma for the entropy together with convexity inequalities. However the bounds are more intricated as in the iid case, and therefore the differential inequalities differ from the optimal differentiel inequality of Bousquet in [1] (2002). Consequently the results are suboptimal in the large deviations bandwith. We start by right-hand side deviations.

Theorem 1 *Let $\mathcal{S}$ be a countable class of measurable functions with values in $[-1,1]^n$. Suppose that $\mathbb{E}(s^i(X_i)) = 0$ for any $s = (s^1, \dots s^n)$ in $\mathcal{S}$ and any integer i in $[1,n]$. Set $V_n = \sup\{\operatorname{Var} S_n(s) : s \in \mathcal{S}\}$. Let L denote the logarithm of the Laplace transform of Z. Then, for any positive t,*

$$L(t) \leq t\mathbb{E}(Z) + \frac{t}{2}(2\mathbb{E}(Z) + V_n)(\exp(e^t - 1) - 1). \tag{2.3}$$

Consequently, for any positive x,

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + x) \leq \exp\Big(- \frac{x}{2} \log\Big(1 + \log(1 + x/(2\mathbb{E}(Z) + V_n))\Big)\Big) \tag{2.4}$$

and

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + \sqrt{2x(V_n + 2\mathbb{E}(Z))} + x) \leq \exp(-x). \tag{2.5}$$

Remark 1 *In the spirit of Massart's paper (see [5]), Theorem 1 (2.4) can be improved for large values of x to get a Bennett's type inequality.*

Remark 2 *Some different concentration inequalities for set-indexed empirical processes are given by Rio in [6] (2001). These concentration can be applied to non identically distributed random variables (cf. Rio [6]) Theorem 4.2 and Remark 4.1). However, due to the concavity of the polynomial function $u(1-u)$, the variance factor was suboptimal for non identically distributed random variables. Here, as a byproduct of Theorem 1 (2.3), we get below the upper bound for the variance of Z.*

Corollary 1 *Under the assumptions of Theorem 1 (2.3),* $\operatorname{Var} Z \leq V_n + 2\mathbb{E}(Z)$.

For left-hand side deviations, the concentration bounds are similar. We emphasize that the proof of Theorem 1 is not relevant for left-hand side deviations. This is the reason why we need to compensate the empirical process for left-hand side deviations. However, due to technical difficulties, we are able to bound up the Laplace tranform only on some finite interval. So, let ψ be the function defined by

$$\psi(t) = (\exp(2t) + 1)/2. \tag{2.6}$$

We denote by t_0 the positive solution of the equation

$$\psi(t) \log \psi(t) = 1. \tag{2.7}$$

The so-defined real t_0 belongs to $[0.463, 0.464]$. Theorem 2 (2.8) below extends Theorem 1 (2.3) to the interval $[-t_0, 0]$. Next we proceed as Klein in [2] (2002) to get a Bennett-type inequality without constraint on x.

Theorem 2 *Under the assumptions of Theorem 1, for any t in $[0, t_0]$,*

$$L(-t) \leq -t\mathbb{E}(Z) + \frac{t}{9}(2\mathbb{E}(Z) + V_n)(\exp(3t) - 3t - 1). \tag{2.8}$$

For any positive x,

$$\mathbb{P}(Z \leq \mathbb{E}(Z) - x) \leq \exp\Big(- \frac{2\mathbb{E}(Z) + V_n}{9} h\Big(\frac{3x}{2\mathbb{E}(Z) + V_n}\Big)\Big), \tag{2.9}$$

where $h(x) = (1 + x)\log(1 + x) - x$.

Remark 3 *Theorem 2 (2.9) improve on Theorem 1.1, inequality (2) in Klein (see [2]). However Klein gives additionnal results for functions with values in $]-\infty, 1]$ and subexponential tails on the left (cf. inequality (3), Theorem 1.1).*

Let us now discuss the applications of Theorems 1 and 2. One of the most popular empirical processes associated to non identically distributed random variables is the so-called Rademacher process, as defined by Ledoux and Talagrand in [4] (1991), section 4.3. Let $(\varepsilon_i)_{i>0}$ be a sequence of iid random variables with $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. Let T be some countable set and $(\varphi_i)_{i>0}$ be a sequence of real-valued functions on T such that $\sum_i (\varphi_i(t))^2 < \infty$ for each t in T. Assume that

$$Z = \sup_{t\in T} \Big| \sum_{i>0} \varepsilon_i \varphi_i(t) \Big| < \infty \text{ a.s.}$$

Then, if

$$V = \sup_{t\in T} \sum_{i>0} (\varphi_i(t))^2 < \infty,$$

the random variable Z satisfies the concentration inequality (4.10) in Ledoux and Talagrand [4] (1991): if m_Z denotes a median of Z, then

$$\mathbb{P}(Z \geq m_Z + 2\sqrt{2Vx}) \leq 4\exp(-x). \tag{2.10}$$

This inequality implies the integrability of Z and some similar inequality for the concentration of Z around it means. If x is not too large, Theorem 1 (2.5) may be used to get a sharper bound. Setting

$$M = \sup_{i>0} \sup_{t \in T} |\varphi_i(t)|,$$

the bounds in Theorem 1 yield:

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + \sqrt{2(2\mathbb{E}(Z) + V)x} + Mx) \leq \exp(-x). \tag{2.11}$$

This new inequality is relevant if $V/M^2 \gg 1$ and $\mathbb{E}(Z) \ll V$. In that case (2.11) improves on (2.10) in the moderate deviations bandwith $x \ll V/M^2$.

2.2 Tensorization of entropy and related inequalities.

In this section we apply the method of tensorization of the entropy to get an upper bound on the entropy of positive functionnals f of independent random variables $X_1, X_2, \ldots, X_n$.

Let $\mathcal{F}_n$ be the σ-field generated by $(X_1, \ldots, X_n)$ and $\mathcal{F}_n^k$ be the σ-field generated by $(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let E_n^k denote the conditional expectation operator associated to $\mathcal{F}_n^k$. Let f be some positive $\mathcal{F}_n$-measurable random variable such that $\mathbb{E}(f \log f) < \infty$. In this paper, the main tool for proving concentration inequalities is the following consequence of the tensorization inequality in Ledoux [3] (1996).

Proposition 1 *Let $g_1, g_2, \ldots, g_n$ be any sequence of positive and integrable random variables such that $\mathbb{E}(g_i \log g_i) < \infty$. Then*

$$\mathbb{E}(f \log f) - \mathbb{E}(f) \log \mathbb{E}(f) \leq \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k / E_n^k g_k)) + \sum_{k=1}^{n} \mathbb{E}((f - g_k) \log(f / E_n^k f)).$$

Proof. Set $f_k = E_n^k f$. By the tensorization inequality in Ledoux (1996)

$$\mathbb{E}(f \log f) \leq \sum_{k=1}^{n} \mathbb{E}(f \log(f/f_k)). \tag{2.12}$$

Now

$$\mathbb{E}(f \log(f/f_k)) = \mathbb{E}(g_k \log(f/f_k)) + \mathbb{E}((f - g_k) \log(f/f_k)).$$

Since $E_n^k(f/f_k) = 1$, we have:

$$\mathbb{E}(g_k \log(f/f_k)) \leq \sup\{\mathbb{E}(g_k h) : \ h \ \mathcal{F}_n\text{-measurable}, \ E_n^k(e^h) = 1\}.$$

Hence, from the duality formula for the relative entropy in Ledoux (1996),

$$\mathbb{E}(g_k \log(f/f_k)) \le \mathbb{E}(g_k \log(g_k/E_n^k g_k)).$$

It follows that

$$\mathbb{E}(f \log(f/f_k)) \le \mathbb{E}(g_k \log(g_k/E_n^k g_k)) + \mathbb{E}((f-g_k)\log(f/f_k)), \tag{2.13}$$

and Proposition 1 follows.

2.3 Right-hand side deviations.

To prove Theorems 1 and 2, we start by proving the results for a finite class of functions. The reults in the countable case are derived from the finite case using the Beppo-Levi lemma. Consequently, throughout the sequel we may assume that $\mathcal{S} = \{s_1, \dots, s_m\}$.

As mentionned in the introduction, the deviation of Z on the right is easier to handle than the deviation on the left. In fact, for positive t, the functionnal $\exp(tZ)$ is an increasing and convex function with respect to the variables $s_i^k(X_k)$. This is no more the case for negative values of t. Consequently, upper bounds for the Laplace transform of Z via the Herbst-Ledoux method are more difficult to handle for negative values of t. In section 2.4, we will introduce compensated processes in order to handle the deviation on the left.

Throughout section 2.3, t is a positive real. Let τ be the first integer such that $Z = S_n(s_\tau)$. Set $f = \exp(tZ)$ and $f_k = E_n^k(f)$. Let P_n^k denote the conditional probability measure conditionally to $\mathcal{F}_n^k$. Set

$$g_k = \sum_i P_n^k(\tau = i)\exp(tS_n(s_i)). \tag{2.14}$$

Let F denote the Laplace transform of Z. From Proposition 1,

$$tF'(t) - F(t)\log F(t) \le \sum_{k=1}^n \mathbb{E}(g_k \log(g_k/E_n^k g_k)) + \sum_{k=1}^n \mathbb{E}((f-g_k)\log(f/E_n^k f)). \tag{2.15}$$

Clearly $g_k \le f$, which implies that

$$\mathbb{E}((f-g_k)\log(f/f_k)) \le \log(\|f/f_k\|_\infty)\mathbb{E}(f-g_k). \tag{2.16}$$

Therefore the upper bound on the second term in (3.3) will be derived from Lemma 2 below.

Lemma 2 *Amost surely* $(f/f_k) \le \exp(ts_\tau^k(X_k)) \le \exp(t)$.

Proof. Let $S_n^k(s) = S_n(s) - s^k(X_k)$. Let τ_k be the first integer in $[1, m]$ such that

$$S_n^k(s_{\tau_k}) = \sup\{S_n^k(s) : s \in \mathcal{S}\}.$$

Clearly

$$f \geq \exp(tS_n^k(s_{\tau_k}))\exp(ts_{\tau_k}^k(X_k)). \tag{2.17}$$

Since the stopping time τ_k is $\mathcal{F}_n^k$-measurable, $E_n^k(s_{\tau_k}^k(X_k)) = 0$ by the centering assumption on the elements of $\mathcal{S}$. It follows that

$$E_n^k f \geq \exp(tS_n^k(s_{\tau_k}))E_n^k(\exp(ts_{\tau_k}^k(X_k)) \geq \exp(tS_n^k(s_{\tau_k})). \tag{2.18}$$

Now $\exp(tS_n^k(s_{\tau_k})) \geq \exp(tS_n^k(s_\tau))$, which ensures that

$$f \leq f_k \exp(ts_\tau^k(X_k)) \leq f_k \exp(t).$$

Hence Lemma 2 holds. ■

From both (2.15), (2.17) and Lemma 2 we get that

$$\mathbb{E}(f\log(f/f_k)) \leq \mathbb{E}(g_k\log(g_k/E_n^k g_k)) + t\mathbb{E}(f - g_k). \tag{2.19}$$

Let

$$h_k = \sum_{i=1}^m P_n^k(\tau = i)\exp(tS_n^k(s_i)). \tag{2.20}$$

The random variable h_k is positive and $\mathcal{F}_n^k$-measurable. Hence

$$\mathbb{E}(g_k\log(g_k/E_n^k g_k)) = \mathbb{E}\Big(h_k E_n^k((g_k/h_k)\log(g_k/E_n^k g_k))\Big).$$

Next, by the Young inequality $xy \leq x\log x - x + e^y$ applied to $x = (g_k/h_k)$ and $e^y = (g_k/E_n^k g_k)$,

$$h_k E_n^k((g_k/h_k)\log(g_k/E_n^k g_k)) \leq E_n^k((g_k\log(g_k/h_k) - g_k + h_k),$$

which ensures that

$$E_n^k(g_k\log(g_k/E_n^k g_k)) \leq E_n^k((g_k\log(g_k/h_k) - g_k + h_k).$$

Putting this inequality in (2.19), we get

$$\mathbb{E}(f\log(f/f_k)) \leq \mathbb{E}(g_k\log(g_k/h_k) + (1+t)(h_k - g_k)) + t\mathbb{E}(f - h_k). \tag{2.21}$$

In order to bound up the second term on right hand, we will use Lemma 3 below.

Lemma 3 *Let $(h_k)_{k\leq n}$ be the finite sequence of random variables defined in (2.20). Then*

$$\sum_{k=1}^n \mathbb{E}(f - h_k) \leq e^t F(t)\log F(t).$$

Proof. Since the random variables $S_n^k(s)$ are $\mathcal{F}_n^k$ measurable,

$$h_k = E_n^k\Big(\sum_{i=1}^m \mathbb{I}_{\tau=i} \exp(tS_n^k(s_i))\Big) = E_n^k\Big(\exp(tS_n^k(s_\tau))\Big).$$

It follows that

$$\sum_{k=1}^n \mathbb{E}(f - h_k) = \sum_{k=1}^n \mathbb{E}\Big(f(1 - \exp(-ts_\tau^k(X_k)) - e^t ts_\tau^k(X_k))\Big) + te^t F'(t). \tag{2.22}$$

Now, from Lemma 2, $ts_\tau^k(X_k) \geq \log(f/f_k)$. Since $1 - \exp(-tx) - te^t x$ is a decreasing function of x on the interval $[-1, 1]$, it follows that

$$\mathbb{E}(f(1 - \exp(-ts_\tau^k(X_k)) - te^t s_\tau^k(X_k))) \leq \mathbb{E}(f - f_k - e^t f \log(f/f_k)).$$

From the equality $\mathbb{E}(f_k) = \mathbb{E}(f)$, we get that

$$\mathbb{E}(f - f_k - e^t f \log(f/h_k)) = -e^t \mathbb{E}(f \log(f/f_k)).$$

Hence, summing on k and applying (2.14),

$$\sum_{k=1}^n \mathbb{E}\Big(f(1 - \exp(-ts_\tau^k(X_k)) - te^t s_\tau^k(X_k))\Big) \leq e^t(F \log F - tF'),$$

which, together with (2.22), implies Lemma 3. ■

Next, we have to bound up the first term on right hand in (2.21). Let

$$r(t, x) = x \log x + (1 + t)(1 - x).$$

Clearly

$$g_k \log(g_k/h_k) + (1 + t)(h_k - g_k)) = h_k r(t, g_k/h_k).$$

From the convexity of r with respect to x,

$$h_k r(t, g_k/h_k) \leq \sum_i P_n^k(\tau = i) \exp(tS_n^k(s_i)) r(t, \exp(ts_i^k(X_k)).$$

which ensures that

$$E_n^k(h_k r(t, g_k/h_k)) \leq \sum_i P_n^k(\tau = i) \exp(tS_n^k(s_i)) \mathbb{E}\Big(r(t, \exp(ts_i^k(X_k))\Big). \tag{2.23}$$

Here we need the bound below.

Lemma 4 *For any function s in $\mathcal{S}$,*

$$\mathbb{E}\, r(t, \exp(ts^k(X_k))) \leq \frac{t^2}{2} \mathbb{E}(s^k(X_k))^2,$$

Proof. Let

$$\eta(x) = r(t, e^{tx}) = txe^{tx} + (t+1)(1 - e^{tx}).$$

We will prove that, for any x in $]-\infty, 1]$,

$$\eta(x) \leq x\eta'(0) + (tx)^2/2. \tag{2.24}$$

Indeed,

$$\eta(x) - x\eta'(0) = tx(e^{tx} - 1) - (1+t)(e^{tx} - 1 - tx).$$

If x is negative, then

$$\eta(x) - x\eta'(0) \leq txe^{tx} - e^{tx} + 1 \leq (tx)^2/2.$$

If x belongs to $[0,1]$, then

$$\eta(x) - x\eta'(0) \leq tx(e^{tx} - 1) - (1+tx)(e^{tx} - 1 - tx) \leq (tx)^2 - (e^{tx} - tx - 1) \leq (tx)^2/2.$$

Hence (2.24) holds. Since the random variables $s^k(X_k)$ are centered, taking $x = s^k(x_k)$ and integrating with respect to the marginal law of X_k, we get Lemma 4. ■

From Lemma 4 and (2.23) we have:

$$E_n^k(h_k r(t, g_k/h_k)) \leq \frac{t^2}{2} E_n^k\Big(\sum_i \mathbb{I}_{\tau=i} \exp(tS_n^k(s_i))\mathbb{E}(s_i^k(X_k))^2\Big). \tag{2.25}$$

Now $\exp(tS_n^k(s_i)) \leq \exp(t + tS_n(s_i))$, and therefrom

$$\sum_{k=1}^n \mathbb{E}(h_k r(t, g_k/h_k)) \leq \frac{t^2 e^t}{2} \mathbb{E}\Big(\sum_i \mathbb{I}_{\tau=i} \exp(tS_n(s_i)) \sum_{k=1}^n \mathbb{E}(s_i^k(X_k))^2\Big).$$

Since $\sum_k \mathbb{E}(s_i^k(X_k))^2 \leq V_n$, we infer that

$$\sum_{k=1}^n \mathbb{E}(h_k r(t, g_k/h_k)) \leq \frac{1}{2} t^2 e^t V_n F(t). \tag{2.26}$$

Together with Lemma 2 and (2.21), (2.26) leads to the differential inequality :

$$tL' - (te^t + 1)L \leq t^2 e^t (V_n/2). \tag{2.27}$$

Let $\gamma(t) = t^{-2}\exp(1 - \exp(t))$. Then

$$\gamma(t)(tL' - (te^t + 1)L) = (t\gamma L)'.$$

Hence, multiplying (2.27) by γ, we get that

$$(t\gamma L)' \leq (V_n/2)e^t \exp(1 - e^t). \tag{2.28}$$

Since $t\gamma(t) \sim (1/t)$ as t tends to 0, integrating (2.28) gives

$$t\gamma(t)L(t) \leq \mathbb{E}(Z) + (V_n/2)(1 - \exp(1 - e^t)),$$

which implies Theorem 1 (2.3). ∎

To prove Theorem 1 (2.4) we apply both Markov's inequality to the random variable $\exp(tZ)$ and Theorem 1 (2.3) with

$$t = \log\Big(1 + \log(1 + x/(2\mathbb{E}(Z) + V_n))\Big).$$

Theorem 1 (2.5) is a straightforward byproduct of the upper bound on the log-Laplace transform of $Z - \mathbb{E}(Z)$ below (cf. Rio (2000), p. 153).

Lemma 5 *Under the assumptions of Theorem 1, for any t in $]0,1[$,*

$$L(t) \leq t\mathbb{E}(Z) + (2\mathbb{E}(Z) + V_n)\frac{t^2}{2 - 2t}.$$

Proof. From Theorem 1 (2.3), it is enough to prove that

$$\exp(e^t - 1) \leq 1 + t/(1 - t).$$

This inequality holds if and only if

$$\lambda(t) = 1 - e^t - \log(1 - t) \geq 0. \tag{2.29}$$

Expanding λ in power series yields $\lambda(t) = \sum_{j>0} b_j t^j$ with $b_j = (1/j) - (1/j!)$. Now (2.29) is an immediate consequence of the fact that $b_j \geq 0$ for any $j > 0$. ∎

2.4 Compensated empirical processes.

In this section we prove Theorem 2. We start by proving Theorem 2 (2.8). Throughout the section, t is any positive real. For i in $\{1, \ldots, m\}$, let

$$L_i(t) = \log \mathbb{E}\Big(\exp(-tS_n(s_i))\Big).$$

Let us define the exponentially compensated empirical process $T(s_i, t)$ by

$$T(s_i, t) = S_n(s_i) + t^{-1}L_i(t). \tag{2.30}$$

We set

$$Z_t = \sup_{1 \leq i \leq m} T(s_i, t) \text{ and } f_t = \exp(-tZ_t). \tag{2.31}$$

Let

$$F(t) = \mathbb{E}(f_t) = \mathbb{E}(\exp(-tZ_t)). \tag{2.32}$$

Our purpose is to obtain a differential inequality for F via the log-Sobolev method. Let τ_t denote the first integer i such that $Z_t = T(s_i, t)$. Since the random functions $T(s_i, t)$ are analytical functions of t, the random function f_t is continuous and piecewise analytic, with derivative with respect to t, almost everywhere (a.e.)

$$f_t' = -Z_t f_t - (L'_{\tau_t}(t) - t^{-1} L_{\tau_t}(t)) f_t = -(Z_t + tZ_t') f_t, \tag{2.33}$$

where $tZ_t' = L'_{\tau_t}(t) - t^{-1} L_{\tau_t}(t)$ by convention. Consequently the Fubini theorem applies and

$$F(t) = 1 - \int_0^t \mathbb{E}((Z_u + uZ_u') f_u) du. \tag{2.34}$$

Therefrom the function F is absolutely continuous with respect to the Lebesgue measure, with a.e. derivative in the sense of Lebesgue

$$F'(t) = -\mathbb{E}((Z_t + tZ_t') f_t). \tag{2.35}$$

Moreover, from the elementary lower bound $f_t \geq \exp(-2nt)$, the function $\log F$ is absolutely continuous with respect to the Lebesgue measure, with a.e. derivative F'/F if F' is the above defined function.

We now apply Proposition 1 to the random function f_t. Clearly

$$\mathbb{E}(f_t \log f_t) - \mathbb{E}(f_t) \log \mathbb{E}(f_t) = \mathbb{E}(t^2 Z_t' f_t) + tF'(t) - F(t) \log F(t) \quad \text{a.e.} \tag{2.36}$$

Hence, applying Proposition 1 with $f = f_t$ and setting $f_k = E_n^k f_t$,

$$tF' - F \log F \leq -\mathbb{E}(t^2 Z_t' f) + \sum_{k=1}^n \mathbb{E}(g_k \log(g_k / E_n^k g_k)) + \sum_{k=1}^n \mathbb{E}((g_k - f)) \log(f_k/f)). \tag{2.37}$$

Now choose

$$g_k = \sum_i P_n^k(\tau_t = i) \exp(-tS_n(s_i) - L_i(t)). \tag{2.38}$$

By definition of Z_t,

$$\exp(-tS_n(s_i) - L_i(t)) \geq \exp(-tZ_t),$$

which implies that $g_k \geq f$. Therefore the upper bound on the second term in (2.37) will be derived from Lemma 6 below.

Lemma 6 *Set $l_{ki}(t) = \log \mathbb{E}(\exp(-ts_i^k(X_k)))$. For fixed t, let $\tau = \tau_t$. Then a.s.*

$$(f_k/f) \leq \exp(ts_\tau^k(X_k) + l_{k\tau}(t)) \leq \psi(t).$$

Proof. Let $S_n^k(s) = S_n(s) - s^k(X_k)$. Set

$$Z^{(k)} = \sup\{S_n^k(s) + t^{-1}\log \mathbb{E}(\exp(-tS_n^k(s))) : s \in \mathcal{S}\}.$$

Let τ_k be the first integer in $[1, m]$ such that

$$S_n^k(s_{\tau_k}) + t^{-1}\log \mathbb{E}(\exp(-tS_n^k(s_{\tau_k})) = Z^{(k)}$$

Clearly

$$f_t \leq \exp(-tZ^{(k)})\exp(-ts_{\tau_k}^k(X_k)) - l_{k\tau_k}(t)).$$

Since the stopping time τ_k is $\mathcal{F}_n^k$-measurable, it follows that

$$E_n^k f_t \leq \exp(-tZ^{(k)}). \tag{2.39}$$

Now, by definition of $Z^{(k)}$,

$$\exp(-tZ^{(k)}) \leq \exp(-tZ + ts_\tau^k(X_k) + l_{k\tau}(t)), \tag{2.40}$$

which ensures that

$$(f_k/f) \leq \exp(ts_\tau^k(X_k) + l_{k\tau}(t)).$$

To conclude the proof of Lemma 6, recall that $\mathbb{E}(\exp(tX)) \leq \cosh(t)$, for any centered random variable X with values in $[-1, 1]$, which implies the second part of Lemma 6. ■

The next step to bound up the second term on right hand is Lemma 7 below.

Lemma 7 *Let (g_k) be the finite sequence of random variables defined in (2.38). Set $\varphi = \psi \log \psi$. Then, for any t in $[0, t_0[$,*

$$\sum_{k=1}^n \mathbb{E}\Big((g_k - f)\log(f_k/f)\Big) \leq \frac{\varphi(t)}{1-\varphi(t)}\left(\sum_{k=1}^n \mathbb{E}(g_k \log(g_k/E_n^k g_k)) - \mathbb{E}(f\log f)\right).$$

Proof. Since the random variables $S_n^k(s)$ are $\mathcal{F}_n^k$ measurable,

$$E_n^k(g_k) = \sum_{i=1}^m P_n^k(\tau = i)\exp(-tS_n^k(s_i) - L_i(t) + l_{ki}(t)). \tag{2.41}$$

It follows that $E_n^k(g_k) = E_n^k(f\exp(ts_\tau^k(X_k) + l_{k\tau}(t)))$. Hence

$$\sum_{k=1}^n \mathbb{E}(g_k - f) = \sum_{k=1}^n \mathbb{E}\Big(f(\exp(ts_\tau^k(X_k) + l_{k\tau}(t)) - 1)\Big). \tag{2.42}$$

Now, for x in $[-1, 1]$, the function $x \to e^x - 1 - x\psi(t)$ is nonincreasing. Now, setting $\eta_k = ts_\tau^k(X_k) + l_{k\tau}(t)$, we have

$$\sum_{k=1}^n \mathbb{E}(g_k - f) = \sum_{k=1}^n \mathbb{E}\Big((e^{\eta_k} - 1 - \eta_k\psi(t) + \eta_k\psi(t))f\Big).$$

Since $\eta_k \geq \log(f_k/f)$ by Lemma 6, we infer that

$$\sum_{k=1}^{n} \mathbb{E}(g_k - f) \leq \sum_{k=1}^{n} \mathbb{E}\Big(f\big((f_k/f) - 1 - \psi(t)\log(f_k/f)\big) + \eta_k f \psi(t)\Big)$$
$$\leq \psi(t) \left(\sum_{k=1}^{n} \mathbb{E}(f \log(f/f_k)) - \mathbb{E}(f \log f) \right).$$

Hence, applying (2.13), we obtain

$$\sum_{k=1}^{n} \mathbb{E}(g_k - f) \leq \psi(t) \left(\sum_{k=1}^{n} \mathbb{E}\Big(g_k \log(g_k/E_n^k g_k) + (g_k - f)\log(f_k/f)\Big) - \mathbb{E}(f \log f) \right). \quad (2.43)$$

Now, from Lemma 6 we know that $\log(f_k/f) \leq \log \psi(t)$. Since $(g_k - f) \geq 0$, it follows that

$$\sum_{k=1}^{n} \mathbb{E}\Big((g_k - f)\log(f_k/f)\Big) \leq \log \psi(t) \sum_{k=1}^{n} \mathbb{E}(g_k - f). \quad (2.44)$$

Putting this bound in (2.43), we infer that

$$(1 - \psi(t)\log\psi(t)) \sum_{k=1}^{n} \mathbb{E}(g_k - f) \leq \psi(t) \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k/E_n^k g_k)) - \psi(t)\mathbb{E}(f \log f).$$

Since $1 - \psi(t)\log\psi(t) > 0$ for any t in $[0, t_0[$, both the above inequality and (2.44) imply lemma 7. ■

From Lemma 7 and the differential inequality (2.37) we then get that

$$(1 - \varphi)(tF' - F log F) \leq \varphi \mathbb{E}(t^2 Z_t' f - f \log f) - \mathbb{E}(t^2 Z_t' f) + \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k/E_n^k g_k)).$$

Now from (2.36) we know that $\mathbb{E}(t^2 Z_t' f - f \log f) = -tF'$, whence

$$tF' - (1 - \varphi)F \log F \leq -\mathbb{E}(t^2 Z_t' f) + \sum_{k=1}^{n} \mathbb{E}(g_k \log(g_k/E_n^k g_k)). \quad (2.45)$$

Let us now bound up the first term on right hand in (2.45). Set $w_k = (g_k/E_n^k g_k)$. Then

$$E_n^k(g_k \log(g_k/E_n^k g_k)) = E_n^k(g_k) E_n^k(w_k \log w_k).$$

From (2.41), by convexity of the function $x \log x$,

$$E_n^k(g_k) w_k \log w_k \leq \sum_i P_n^k(\tau = i)(-ts_i(X_k) - l_{ki}(t)) \exp(-tS_n(s_i) - L_i(t)).$$

Consequently

$$E_n^k(g_k \log(g_k/E_n^k g_k)) \leq \sum_i P_n^k(\tau = i) \exp(-tS_n^k(s_i) - L_i(t) + l_{ki}(t))(tl'_{ki}(t) - l_{ki}(t)).$$

Since

$$\sum_i P_n^k(\tau = i) \exp(-tS_n^k(s_i) - L_i + l_{ki})(tl'_{ki} - l_{ki}) = E_n^k\Big(\sum_i \mathbb{I}_{\tau=i} \exp(-tS_n^k(s_i) - L_i + l_{ki})(tl'_{ki} - l_{ki})\Big),$$

it implies that

$$\mathbb{E}(g_k \log(g_k/E_n^k g_k)) \leq \mathbb{E}\Big(\exp(-tZ_t + ts_\tau^k(X_k) + l_{k\tau})(tl'_{k\tau} - l_{k\tau})\Big),$$

From the convexity of the functions l_{ki}, we know that $tl'_{k\tau} - l_{k\tau} \geq 0$. Hence, applying Lemma 6, we get

$$\mathbb{E}(g_k \log(g_k/E_n^k g_k))) \leq \psi(t)\mathbb{E}((tl'_{k\tau} - l_{k\tau})f).$$

Since $t^2 Z'_t = tL'_\tau - L_\tau$, it follows that

$$-\mathbb{E}(t^2 Z'_t f) + \sum_{k=1}^n \mathbb{E}(g_k \log(g_k/E_n^k g_k)) \leq (\psi(t) - 1)\mathbb{E}((tL'_\tau - L_\tau)f). \tag{2.46}$$

Both (2.45) and (2.46) yield, for t in $[0, t_0[$,

$$tF' - (1 - \varphi)F \log F \leq (\psi(t) - 1)\mathbb{E}((tL'_\tau - L_\tau)f). \tag{2.47}$$

Set $\Lambda = \log F$. Since $tL'_\tau - L_\tau \leq \sup_i(tL'_i - L_i)$, dividing by F, we infer that

$$t\Lambda' - (1 - \varphi)\Lambda \leq (\psi(t) - 1) \sup_i(tL'_i - L_i). \tag{2.48}$$

Next, we derive an upper bound on $tL'_i - L_i$ from the lemma below.

Lemma 8 *Let Y be a random variable with values in $[-1, 1]$, such that $\mathbb{E}(Y) = 0$ and $\mathbb{E}(Y^2) = v$. Then, for any t in $]0, 2]$,*

$$\mathbb{E}(tYe^{tY}) - \mathbb{E}(e^{tY}) \log \mathbb{E}(e^{tY}) \leq v(1 + (t - 1)e^t).$$

Proof. From the duality formula on the entropy, we know that, for any positive c,

$$H(T) \leq \mathbb{E}(T \log(T/c) - T + c).$$

Taking $T = \exp(tY)$ we then get that

$$H(T) \leq c - 1 - \log c \sum_{k=2}^{\infty} \mathbb{E}(Y^k)\frac{t^k}{k!}(k - 1 - \log c).$$

From the assumptions on Y, we know that $\mathbb{E}(Y^k) \leq v$ for $k \geq 3$. Hence, for $c \leq e^2$,

$$H(T) \leq c - 1 - \log c + v \sum_{k=2}^{\infty} \frac{t^k}{k!}(k - 1 - \log c).$$

Next taking the optimal value $c = 1 + v(e^t - t - 1)$ (note that $c \leq e^2$ for t in $]0, 2]$) in the above upper bound, we obtain

$$H(T) \leq vt(e^t - 1) - (1 + v(e^t - t - 1)) \log(1 + v(e^t - t - 1)) \leq v(1 + (t - 1)e^t),$$

which completes the proof of Lemma 8. ■

Since $tL'_i - L_i = \sum_{k=1}^{n}(tl'_{ki} - l_{ki})$, from Lemma 8 applied to the r.v.'s $s_i^k(X_k)$,

$$tL'_i - L_i \leq V_n(1 + (t - 1)e^t) \tag{2.49}$$

Both the above bound and (2.48) lead to the differential inequality below.

Proposition 2 *For any t in $[0, t_0[$,*

$$t\Lambda' - (1 - \psi \log \psi)\Lambda \leq \tfrac{1}{2} V_n(e^{2t} - 1)(1 + (t - 1)e^t).$$

Before integrating the above differential inequality to get bounds on Λ, we link the log-Laplace L_{-Z} of $-Z$ with Λ.

Lemma 9 *For any positive t,*

$$L_{-Z}(t) - \sup_i L_i(t) \leq \Lambda(t) \leq \min(L_{-Z}(t), 0).$$

Proof. By definition of Z_t,

$$\exp(-tZ_t) = \exp\left(\inf_i(-tS_n(s_i) - L_i(t))\right) \geq \exp(-tZ - \sup_i L_i),$$

Consequently, for any positive t,

$$\exp(\Lambda(t)) \geq \exp(-\sup_i L_i(t)) \mathbb{E} \exp(-tZ),$$

which gives the first inequality. Next, by definition of Z_t,

$$\exp(\Lambda(t)) = \mathbb{E}\Big(\inf_i \exp(-tS_n(s_i) - L_i(t))\Big) \leq \mathbb{E}\Big(\exp(-tS_n(s_1) - L_1(t))\Big) = 1,$$

which ensures that $\Lambda(t) \leq 0$. Moreover $L_i(t) \geq 0$ by the centering assumption on the random variables $S_n(s)$. Hence

$$\exp(\Lambda(t)) \leq \mathbb{E}\Big(\inf_i \exp(-tS_n(s_i))\Big) = \mathbb{E}(\exp(-tZ)),$$

which completes the proof. ■

It remains to bound up Λ. Set $\widetilde{\Lambda}(t) = t^{-1}\Lambda(t)$ and

$$I(t) = \int_0^t \frac{\varphi(u)}{u} du.$$

Then $\left(\widetilde{\Lambda}e^I\right)' = t^{-2}(t\Lambda' - (1-\varphi)\Lambda)e^I$. Consequently, from Proposition 2,

$$\left(\widetilde{\Lambda}e^I\right)' \leq \frac{V_n}{2t^2}(e^{2t}-1)(1+(t-1)e^t)e^h. \tag{2.50}$$

Since $\widetilde{\Lambda}e^I$ is absolutely continuous with respect to the Lebesgue measure, integrating (2.50) yields

$$\widetilde{\Lambda}(t) \leq \widetilde{\Lambda}(\varepsilon)e^{I(\varepsilon)-I(t)} + \frac{V_n}{2}\int_\varepsilon^t u^{-2}(e^{2u}-1)(1+(u-1)e^u)e^{I(u)-I(t)}du \tag{2.51}$$

for $0 < \varepsilon < t$. The control of the integral on right hand side will be done via the bounds for φ below, whose proof is carried out in section 2.5.

Lemma 10 *For any positive t, $t \leq \varphi(t) \leq t\exp(2t) - (t^2/2)$.*

By Lemma 10, $\lim_0 I(\varepsilon) = 0$. Furthermore $\widetilde{\Lambda}(\varepsilon) \leq \varepsilon^{-1}L_{-Z}(\varepsilon)$ by Lemma 9. Therefore

$$\limsup_{\varepsilon\to 0} \widetilde{\Lambda}(\varepsilon)e^{I(\varepsilon)-I(t)} \leq -\mathbb{E}(Z)e^{-I(t)}. \tag{2.52}$$

Now $I(u)-I(t) \leq (u-t)$ by Lemma 10. Consequently, letting $\varepsilon \to 0$ in (2.51) and applying (2.52), we get

$$\Lambda(t) \leq -\mathbb{E}(Z)te^{-I(t)} + \tfrac{1}{2}V_n te^{-t}\int_0^t u^{-2}(e^{2u}-1)(1+(u-1)e^u)e^u du. \tag{2.53}$$

To bound up L_{-Z}, we then apply the Bennett bound $L_i(t) \leq V_n(e^t - t - 1)$ together with Lemma 9. This yields the proposition below.

Proposition 3 *Let the function J be defined by*

$$J(t) = \frac{1}{2}\int_0^t u^{-2}(e^{2u}-1)(1+(u-1)e^u)e^u du.$$

For any t in $[0, t_0[$,

$$L_{-Z}(t) + t\mathbb{E}(Z) \leq t\mathbb{E}(Z)(1-e^{-I(t)}) + V_n te^{-t}J(t).$$

To complete the proof of Theorem 2 (2.8), we bound up the functions appearing in Proposition 3 via Lemma 11 below, proved in section 2.5.

Lemma 11 *For any t in $[0, t_0]$,*

$$(a) \qquad e^t - t - 1 + te^{-t}J(t) \le \tfrac{1}{9}(\exp(3t) - 3t - 1),$$
$$(b) \qquad 1 - \exp(-I(t)) \le \tfrac{2}{9}(\exp(3t) - 3t - 1).$$

Next, proceeding as in Klein ([2]), we prove Theorem 2 (2.9). For sake of brievity, set $E = \mathbb{E}(Z)$. From Theorem 2 (2.8) and Markov's exponential inequality,

$$\mathbb{P}(Z \le E - x) \le \exp\left(\tfrac{1}{9}v(e^{3t} - 3t - 1) - tx\right) \tag{2.54}$$

for any t in $[0, t_0]$ (here $v = 2E + V_n$). Let then

$$x_0 = \tfrac{1}{3}v(\exp(3t_0) - 1). \tag{2.55}$$

For x in $[0, x_0]$, let $t(x)$ be the unique solution of the equation $x = v(\exp(3t) - 1)/3$. The number $t(x)$ belongs to $[0, t_0]$. Consequently, from (2.54) applied with $t = t(x)$ we obtain Theorem 2 (2.8) for x in $[0, x_0]$.

It remains to prove Theorem 1 (2.9) for $x \ge x_0$. Notice then that $x_0 \ge v$. Hence it is enough to prove that Theorem 1 (2.9) still holds for $x \ge v$. Since $Z \ge S_n(s)$ for each s in $\mathcal{S}$, from the Bennett inequality for independent and bounded random variables,

$$\mathbb{P}(Z \le E - x) \le \mathbb{P}(S_n(s) \le x - E) \le \exp(-V_n h((x - E)/V_n)) \tag{2.56}$$

for $x \ge \mathbb{E}(Z)$. To complete the proof, we bound up the function on right hand in (2.56) via Lemma 12 below.

Lemma 12 *Let $E = \mathbb{E}(Z)$ and $v = V_n + 2E$. For any $x \ge v$,*

$$9V_n h((x - E)/V_n) > vh(3x/v).$$

Proof. Set $y = x - v$. Then Lemma 12 holds if

$$\Delta(E, y) = 9V_n h(1 + (E + y)/V_n) - (V_n + 2E)h(3 + 3y/(V_n + 2E)) > 0 \tag{2.57}$$

for any $y \ge 0$. Now $\Delta(0, 0) = V_n(9h(1) - h(3)) > 0$ and, noting that $h'(u) = \log(1 + u)$, we have

$$\frac{\partial \Delta}{\partial E}(E, 0) = 9\log(2 + E/V_n) - 2h(3) \ge 9\log 2 - 2h(3) > 0,$$

which ensures that $\Delta(E, 0) > 0$. Next

$$\frac{\partial \Delta}{\partial y}(E, 0) = 9\log(2 + E/V_n) - 3\log 4 > 0$$

and, noting that $h''(u) = 1/(1 + u)$,

$$\frac{1}{9}\frac{\partial^2 \Delta}{\partial y^2}(E, y) = \frac{1}{2V_n + E + y} - \frac{1}{4(V_n + 2E) + 3y} > 0.$$

Consequently $\Delta(E, y)$ is a convex and increasing function of y. Since $\Delta(E, 0) > 0$, it implies 2.57, and therefrom Lemma 12.

2.5 Technical tools.

In this section, we prove Lemmas 10 and 11.

Proof of Lemma 10. By definition of ψ and φ,

$$\varphi(t) = t\psi(t) + \psi(t)\log\cosh(t) \geq t,$$

since $\psi(t) \geq 1$ for any nonnegative t. Next

$$t\exp(2t) - (t^2/2) - \varphi(t) = \psi(t)\big(t\tanh(t) - \log\cosh(t) - (e^{2t}+1)^{-1}t^2\big),$$

so that Lemma 10 holds if

$$p(t) := t\tanh(t) - \log\cosh(t) - (e^{2t}+1)^{-1}t^2 \geq 0$$

for t in $[0, t_0]$. Now $p(0) = 0$ and

$$2\cosh^2(t)p'(t) = 2t - t(e^{-2t}+1) - t^2 = t(1 - t - e^{-2t}).$$

Since $\exp(-2t) \leq 1 - t$ for t in $[0, 1/2]$, the above identity ensures that $p'(t) \geq 0$ on $[0, t_0]$, which implies Lemma 10.

Proof of Lemma 11. We start by proving (a). Clearly (a) holds if

$$\alpha(t) = \tfrac{1}{9}e^t(e^{3t} - 3t - 1) + e^t(1 + t - e^t) - J(t) \geq 0 \tag{2.58}$$

for any t in $[0, 4]$. (with the convention $\alpha(0) = 0$). The function α is analytic on the real line. To prove (2.58), we then note that $\alpha^{(i)}(0) = 0$ for $i = 1, 2$. Consequently (a) holds if, for t in $[0, 4]$,

$$\alpha^{(3)}(t) = e^{3t}(-t + (19/3)) - 4e^{2t} + e^t(-3t - 5) + (8/3) > 0. \tag{2.59}$$

Now (2.59) holds if $\alpha^{(4)}(t) > 0$, since $\alpha^{(3)}(0) > 0$. Next

$$\beta(t) := e^{-t}\alpha^{(4)}(t) = 3e^{2t}(-2t + 11) - 8e^t - 3$$

satisfies $\beta(0) > 0$ and, for t in $[0, 4]$,

$$\beta'(t) = 12e^{2t}(5 - t) - 8e^t > e^t(12e^t - 8) > 0,$$

which ensures that $\beta(t) > 0$ for t in $[0, 4]$. Hence Lemma 11 (a) holds.

To prove (b), we apply Lemma 10 to bound up the function $I(t)$. This gives

$$I(t) \leq \int_0^t (e^{2u} - u/2)du = \frac{e^{2t} - 1}{2} - \frac{t^2}{4}.$$

Now, recall $t_0 \leq 1/2$. For t in $[0, 1/2]$, expanding $\exp(2t)$ in entire series yields

$$(\exp(2t) - 1)/2 = t + t^2 + 4t^3 \sum_{k\geq 3} \frac{1}{k!}(2t)^{k-3} \leq t + t^2 + 4t^3 \sum_{k\geq 3} \frac{1}{k!}.$$

Hence, for $t \leq t_0$,

$$I(t) \leq t + \tfrac{3}{4}t^2 + (4e - 10)t^3 \leq t + \tfrac{3}{4}t^2 + \tfrac{7}{8}t^3 := \gamma(t). \tag{2.60}$$

From (2.60), Lemma 11 b) holds if

$$d(t) = \tfrac{2}{9}(e^{3t} - 3t - 1) - t + t\exp(-\gamma(t)) \geq 0.$$

Now $d(0) = d'(0) = 0$ and

$$d''(t) = 2e^{-\gamma(t)}\Big(e^{4t+(3/4)t^2/4+(7/8)t^3} - 1 - \tfrac{7}{4}t - \tfrac{15}{4}t^2 + \tfrac{15}{4}t^3 + \tfrac{63}{16}t^4 + \tfrac{441}{128}t^5\Big).$$

Now

$$e^{4t+(3/4)t^2/4+(7/8)t^3} \geq e^{4t} \geq 1 + 4t + 8t^2,$$

which ensures that $d''(t) > 0$ for any positive t. Consequently $d(t) \geq 0$, which implies Lemma 11 b).

Bibliography

[1] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. *To appear in the proceedings of the conference on stochastic inequalities. Eds Giné, Houdré and Nualart.* 2002.

[2] T. Klein. Une inégalité de concentration à gauche pour les processus empiriques. *C.R. Acad. Sci. Paris, Série I,* ***334****, 495-500.* 2002.

[3] M. Ledoux. On Talagrand's deviation inequalities for product measures. *European series in applied and industrial mathematics, Probability and statistics* ***1****, 63-87.* 1996.

[4] M.Ledoux, M. Talagrand. Probability in Banach spaces. *Isoperimetry and Processes. Berlin: Springer.* 1991.

[5] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Annals of Probab.* ***28****, 863-884.* 2000.

[6] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Relat. Fields* ***119****, 163-175.* 2001.

[7] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques *Annales de l'institut Henri Poincaré, probabilités et statistiques* ***38****, 1053-1057.* 2002.

[8] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.* ***126****, 503-563.* 1996.

Part III

Inégalités de concentration convexe

Chapter 3

Cas des variables négativement associées

Convex concentration inequalities for nondecreasing processes

Thierry KLEIN
Université de versailles-Saint-Quentin
45 avenue des Etats-Unis
78035 Versailles cedex
klein@math.uvsq.fr

September the 15th 2003

Abstract 1 *In this paper, we prove convex concentration inequalities for discrete and continuous time counting processes. Then we apply these inequalities to prove that the supremum of independent binomial random variables and the supremun of independent Poisson random variables satisfy convex concentration inequalities.*

AMS Classification: 60E15, 60F10.
Keywords: Martingales, counting processes, records, convex concentration inequality, Negatively associated variables, 3−ary search trees.

3.1 Introduction.

In this paper, we will introduce the concept of binomial (resp. Poissonian) convex concentration inequality for discrete (resp. continuous) time process see Definition 1 (resp. Definition 2). We will then give examples of processes who satisfy these inequalities.

This concept was first introduced by Hoeffding in [6]. In this paper, Hoeffding compares $\mathbb{E}\left(\phi(S_n)\right)$ with $\mathbb{E}\left(\phi(S_n^\star)\right)$, when $S_n = \sum_{i=1}^n X_i$ is the sum of independent Bernoulli distributed random variables with parameters p_i and $S_n^\star$ is $B(n,\bar{p})-$distributed with $\bar{p}$ the arithmetic mean of the p_i's.

Proposition 1 *(Hoeffding [6], Shorack-Wellner [12]). Let $b_1, \ldots, b_n$ be independent random variables Bernoulli distributed with parameters p_i and $S_n = b_1 + \cdots + b_n$. Let $\bar{p} = \dfrac{p_1 + \ldots + p_n}{n}$ then for any convex function ϕ we have*

$$\mathbb{E}\Big(\phi(S_n)\Big) \leq \mathbb{E}\Big(\phi\big(B(n,\bar{p})\big)\Big). \tag{3.1}$$

These inequalities are very useful to derive tail inequalities as pointed by Hoeffding [6], Bretagnolle [3] who gave a functional version of this result, Pinelis in [7] and [8] studies a more general case where the function ϕ is in a general class of functions. Shao in [11] treats the case of Negatively Associated (N.A) random variables and shows how convex concentrations inequalities lead to classical inequalities like Rosenthal maximal inequality or Kolmogorov inequality. In particular, he is able to extend Hoeffding's inequality on the probability bounds for the sum of a random sample without replacement from a finite population. Bentkus in [1] uses convex concentration inequalities to give bounds for probabilities tails of discrete martingales with bounded jumps.

In this paper, we introduce a class of discrete processes which satisfy convex concentration inequalities. Our approach is similar to Shao's approach [11]. Indeed our first result (Theorem 1) states that under some appropriate hypothesis on the discrete process $(Z_n)_{n\in\mathbb{N}}$ (Assumption 1), for any convex function ϕ,

$$\mathbb{E}\left(\phi(Z_n)\right) \leq \mathbb{E}\left(\phi(S_n)\right), \tag{3.2}$$

where S_n is $B(nl, \mathbb{E}(Z_n)/nl)-$distributed[1]. The key argument in the proof of this result is that $(Z_{n+1} - Z_n, Z_n)$ is N.A. for any $n \in \mathbb{N}$.

Next, we give an analogue of Theorem 1 for continuous time counting processes $(A_t)_{t\geq 0}$ (see Dellacherie-Meyer [5] or Brémaud [2] for complete study of these processes and in particular for properties of their compensators). Our result for continuous time process states that under some appropriate hypothesis (Assumption 2 and Assumption 3), for any convex function ϕ

$$\mathbb{E}\left(\phi(A_t)\right) \leq \mathbb{E}\left(\phi(Y_t)\right), \tag{3.3}$$

where Y_t is $\mathcal{P}(\mathbb{E}(A_t))-$distributed (Assumption 2, concerning the absolutely continuity of the compensator of A_t is due to Reynaud-Bourret [10]). The proof of the continuous time theorem (Theorem 2) relies on differential equations.

Section 3.4 is devoted to applications of Theorems 1 and 2. First, we will prove that suprema of binomial (resp. Poissonian) independent random variables are more concentrated in the sence of convex concentration inequality than a single binomial (resp. Poissonian) variable. In other words, let $(Y_i)_{1\leq i\leq p}$ be independent random variables with distribution $B(n, p_i)$ and $Z_n = \sup(Y_1, \ldots, Y_p)$. For any convex function ϕ

$$\mathbb{E}\left(\phi(Z_n)\right) \leq \mathbb{E}\left(\phi(S_n)\right), \tag{3.4}$$

where S_n is $B(n, \mathbb{E}(Z_n)/n)-$distributed. In the same way, if $\left(N^{(i)}\right)_{1\leq i\leq p}$ are independent random variables with Poisson distribution with parameter μ_i and $A_t = \sup(N^{(1)}, \ldots, N^{(p)})$ for any convex function ϕ

$$\mathbb{E}\left(\phi(A_t)\right) \leq \mathbb{E}\left(\phi(Y_t)\right), \tag{3.5}$$

where Y_t is $\mathcal{P}(\mathbb{E}(A_t))-$distributed. The key argument here, is that we are able to compute the compensator of $(A_t)_{t\geq 0}$. The result is, in fact, a concentration inequality for the supremum of a set indexed Poisson process, when the class of sets is a class of disjoint sets. So it is quite natural to formulate the question below.

(Q) Does the process $\left(\sup(\Pi_t(A),\ A\in\mathcal{A})\right)_{t\geq 0}$ satisfy a convex Poissonian concentration inequality when $(\Pi_t)_{t\geq 0}$ is a Poisson process ?

[1] $B(n,p)$ is the binomial distribution with parameters n and p, $\mathcal{E}(\lambda)$ is the exponential distribution of parameter λ, $b(p)$ is the Bernoulli distribution of parameter p and $\mathcal{P}(\mu)$ is the Poisson distribution of parameter μ.

Reynaud-Bourret (see [9]) proved that the answer is positive if we restrict the functions ϕ to be of the form $\phi_\lambda(x) = \exp(\lambda x)$. We can then conjecture that the question (Q) has a positive answer.

In the last application we study the example of 3−ary search trees, and we show that they are an example for which Theorem 1 is valid.

3.2 Definitions and statement of results

Let $(Z_n)_{n\in\mathbb{N}^\star}$ be a nondecreasing discrete time process, with $Z_0 = 0$ and with jumps equal to 1. In this paper we are interested in concentration inequalities for the process Z.

Definition 1 . *A process $(X_n)_{n\in\mathbb{N}}$ is said to satisfy a binomial convex concentration inequality if for any $n \in \mathbb{N}$, any convex function ϕ we have*

$$\mathbb{E}\left(\phi(X_n)\right) \leq \mathbb{E}\left(\phi(Y_n)\right), \tag{3.6}$$

where Y_n is $B(n, \mathbb{E}(X_n)/n)-$distributed.

We will also consider continuous time counting processes. We recall that $(A_t)_{t\geq 0}$ is a counting process if it is a random increasing piecewise constant function with $A_0 = 0$ and with jumps equal to 1 (for a complete description of these processes see Brémaud [2]). Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration and assume that $(A_t)_{t\geq 0}$ is $(\mathcal{F}_t)-$measurable. Let $(\Lambda_t)_{t\geq 0}$ be the compensator of the counting process A, i.e. the nondecreasing function such that $(M_t = A_t - \Lambda_t)_{t\geq 0}$ is a martingale (see Brémaud [2] or Dellacherie and Meyer [5] for a complete description of compensators). In the sequel we are interested in concentration inequalities for the process A.

Definition 2 . *A process $(X_t)_{t\geq 0}$ is said to satisfy a Poissonian convex concentration inequality, if for any $t \geq 0$ and any convex function ϕ we have*

$$\mathbb{E}\left(\phi(X_t)\right) \leq \mathbb{E}\left(\phi(Y_t)\right), \tag{3.7}$$

where Y_t is $\mathcal{P}(\mathbb{E}(X_t))-$distributed.

3.2.1 Theorem for discrete time processes

Let $(Z_n)_{n\in\mathbb{N}^\star}$ be a nondecreasing discrete time process, with $Z_0 = 0$ and jumps equal to 1, *i.e.* $Z_{n+1} - Z_n = 0$ or $Z_{n+1} - Z_n = 1$. We suppose that $(Z_n)_{n\in\mathbb{N}^\star}$ satisfies the following assumption

Assumption 1 . *For any fixed n, the sequence*

$$\left(\mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k)\right)_{k\geq 0}$$

is nonincreasing.

Theorem 1 *(Discrete time). Under Assumption 1, the process $(Z_n)_{n\in\mathbb{N}}$ satisfies a binomial convex concentration inequality. In other words, for any convex function ϕ we have*

$$\mathbb{E}\left(\phi(Z_n)\right) \leq \mathbb{E}\left(\phi(Y_n)\right), \tag{3.8}$$

where Y_n is $B(n, \mathbb{E}(Z_n)/n)-$distributed.

3.2.2 Theorem for continuous time processes

Let $(A_t)_{t\geq 0}$ be a counting process, whose compensator $(\Lambda_t)_{t\geq 0}$ satisfies the following two assumptions.

Assumption 2 . *The compensator $(\Lambda_t)_{t\geq 0}$ is absolutely continuous and a.s. finite on $[0,T]$.*

Note that Assumption 2 implies that A has a.s. a finite number of jumps (recall the jumps are equal to 1). In the sequel we will denote by λ_s the derivative of $d\Lambda_s$ with respect to ds (see Reynaud-Bourret [10] who introduces this assumption and gives other applications for counting processes).

Assumption 3 . $\mathbb{E}(\lambda_t \mid A_{t-})$ *is a nonincreasing function of A_{t-}.*

Theorem 2 *(Continuous time). Under Assumptions 2 and 3, the process $(A_t)_{t\geq 0}$ satisfies a Poissonian convex concentration inequality. In other words, for any convex function ϕ we have*

$$\mathbb{E}\left(\phi(A_t)\right) \leq \mathbb{E}\left(\phi(Y_t)\right), \tag{3.9}$$

where Y_t is $\mathcal{P}(\mathbb{E}(A_t))-$distributed.

3.3 Proofs

3.3.1 Proof of theorem 1

Theorem 1 will be a consequence of Theorem 3, which is Theorem 1 in Shao [11]. We briefly recall Shao's setting. A finite family of random variables $\{X_i,\ 1\leq i\leq n\}$ is said to be negatively associated ($N.A.$) if for every pair of disjoint subsets A_1 and A_2 of $\{1,2,\ldots,n\}$,

$$\mathrm{Cov}\{f_1(X_i, i\in A_1), f_2(X_j, j\in A_2)\} \leq 0, \tag{3.10}$$

whenever f_1 and f_2 are coordinatewise increasing and the covariance exists. An infinite family is N.A. if every finite subfamily is N.A.

Theorem 3 *(Shao [11]). Let $\{X_i, 1\leq i\leq n\}$ be a N.A sequence and let $\{X_i^\star, 1\leq i\leq n\}$ be a sequence of independent random variables such that X_i and $X_i^\star$ have the same distribution for each $i=1,2,\ldots,n$. Then*

$$\mathbb{E}\Big(f\big(\sum_{i=1}^n X_i\big)\Big) \leq \mathbb{E}\Big(f\big(\sum_{i=1}^n X_i^\star\big)\Big) \tag{3.11}$$

for any convex function f on $\mathbb{R}$, whenever the expectation on the right hand side of (3.11) exists.

Remark 1 . *The proof of Theorem 3 requires only that (S_n, X_{n+1}) is N.A. for any $n\in\mathbb{N}$.*

Theorem 3 implies the following lemma.

Lemma 1 . *Let ϕ be a convex function. Under the assumptions of Theorem 1, we have*

$$\mathbb{E}(\phi(Z_n)) \leq \mathbb{E}(\phi(S_n)), \tag{3.12}$$

where S_n is the sum of independent Bernoulli variables, $a_1 + \cdots + a_n$, such that $\mathbb{E}(a_i) = \mathbb{E}\left(Z_i - Z_{i-1}\right)$.

Proof: Let $b_{n+1} = Z_{n+1} - Z_n$ and let us prove that (Z_n, b_{n+1}) is N.A for any $n \in \mathbb{N}$.
We will first prove that the function

$$p(n,t) = \mathbb{P}(b_{n+1} = 1 \mid Z_n \geq t)$$

in nonincreasing in $t \in \mathbb{N}$.
Set $u_n(k) = \mathbb{P}(b_{n+1} = 1 \mid Z_n = k)$ and $\alpha_n(k) = \mathbb{P}(Z_n = k)$, we have

$$p(n,k) = \frac{\sum_{i=0}^{k} u_n(i)\alpha_n(i)}{\sum_{i=0}^{k} \alpha_n(i)},$$

Hence

$$p(n,k) - p(n,k+1) = \frac{\alpha_n(k+1)\sum_{i=0}^{k} u_n(i)\alpha_n(i) - u_n(k+1)\alpha_n(k+1)\sum_{i=0}^{k}\alpha_n(i)}{\sum_{i=0}^{k}\alpha_n(i)\sum_{i=0}^{k+1}\alpha_n(i)}.$$

Using Assumption 1, we get

$$p(n,k) - p(n,k+1) \geq 0.$$

Hence, $p(n,t)$ is nonincreasing.
Let $t \geq 0$, since $p(n,t)$ is nonincreasing we have

$$\mathbb{P}(b_{n+1} = 1 \mid Z_n \geq t) \leq \mathbb{P}(b_{n+1} = 1).$$

This equation can be written as

$$\mathbb{P}(Z_n \geq t,\ b_{n+1} = 1) \leq \mathbb{P}(Z_n \geq t)\mathbb{P}(b_{n+1} = 1).$$

Since $b_{n+1} = 0$ or $b_{n+1} = 1$ we get, for any $(s,t) \in \mathbb{R}^2$,

$$\mathbb{P}(Z_n \geq t,\ b_{n+1} \geq s) \leq \mathbb{P}(Z_n \geq t)\mathbb{P}(b_{n+1} \geq s).$$

In other words

$$\mathrm{Cov}\{I_{Z_n \geq t}\ I_{b_{n+1} \geq s}\} \leq 0. \tag{3.13}$$

From this inequality we get that, for any nondecreasing functions f and g,

$$\mathrm{Cov}\{f(Z_n), g(b_{n+1})\} \leq 0. \tag{3.14}$$

From (3.14) and Theorem 3 (cf. Remark 1), we then get Lemma 1. □
Theorem 1 is an easy consequence of both Lemma 1 and Proposition 1 stated in the introduction.

3.3.2 Proof of theorem 2

We will use differential equation technics to prove Theorem 2. The key point is the lemma below which gives a concrete description of the compensator of $\Big(\phi(A_t)-\phi(A_0)\Big)_{t\geq 0}$.

Lemma 2 . *Let ϕ be a nondecreasing convex function. Then the predictable compensator $\Big(\Lambda_t(A_t,\phi)\Big)_{t\geq 0}$ of $\Big(\phi(A_t)-\phi(A_0)\Big)_{t\geq 0}$ is defined by*

$$\Lambda_t(A_t,\phi)=\int_0^t\Big(\phi(1+A_{s-})-\phi(A_{s-})\Big)\lambda_s ds. \tag{3.15}$$

Proof: Using the fact that the process $(A_t)_{t\geq 0}$ is piecewise constant with jumps equal to 1, we have

$$\phi(A_t)-\phi(A_0)=\int_0^t\Big(\phi(1+A_{s-})-\phi(A_{s-})\Big)dA_s.$$

since the process $(A_t)_{t\geq 0}$ is càdlàg, the process $(A_{t-})_{t\geq 0}$ is left-continuous and so is the process $(\phi(1+A_{s-})-\phi(A_{s-}))_{s\geq 0}$. Using Theorem T8, p. 27 in Brémaud [2] we get that

$$Y_t=\int_0^t\Big(\phi(1+A_{s-})-\phi(A_{s-})\Big)(dA_s-\lambda_s ds)$$

is a $(\mathcal{F}_t)-$martingale. This ends the proof of Lemma 2. □

In order to prove Theorem 2, we will exhibit differential equations satisfied by $\mathbb{E}\left(\phi(A_t)\right)$ and $\mathbb{E}\left(\phi(N_t)\right)$. Denote by $\mathcal{C}$ the set of all convex functions and by $\mathcal{C}_2$ the set of all convex functions of the class C^2. Let $(N_t)_{t\geq 0}$ be a Poisson point process on $\mathbb{R}^+$, with $\mathbb{E}(N_t)=\mathbb{E}(A_t)$. Let

$$h(\phi,t)=\mathbb{E}\left(\phi(A_t)\right),\ g(\phi,t)=\mathbb{E}\left(\phi(N_t)\right).$$

For $a\in\mathbb{R}$, set $A_t^a=A_t+a,\ N_t^a=N_t+a$. Let

$$h_a(\phi,t)=\mathbb{E}\left(\phi(A_t^a)\right),\ g_a(\phi,t)=\mathbb{E}\left(\phi(N_t^a)\right).$$

Note that $h_0=h$ and $g_0=g$. Then

$$g(\phi,t)=\sum_{k=0}^{\infty}\phi(k)e^{\mathbb{E}(A_t)}\mathbb{E}(A_t)^k/k!.$$

Using the definition of λ_t and Fubini's theorem, we get

$$\frac{d}{dt}\mathbb{E}(A_t)=\mathbb{E}(\lambda_t).$$

Consequently

$$\frac{dg}{dt}(\phi,t)=\mathbb{E}(\lambda_t)\Big(-\sum_{k=0}^{\infty}\phi(k)e^{\mathbb{E}(A_t)}\mathbb{E}(A_t)^k/k!+\sum_{k=0}^{\infty}\phi(k+1)e^{\mathbb{E}(A_t)}\mathbb{E}(A_t)^k/k!\Big).$$

This equation can be written in the following way

$$\frac{dg}{dt}(\phi,t) = \mathbb{E}(\lambda_t)\Big(\mathbb{E}(\phi(N_t+1)-\phi(N_t)\Big),$$

whence

$$\frac{dg_0}{dt}(\phi,t) + \mathbb{E}(\lambda_t)g_0(\phi,t) - \mathbb{E}(\lambda_t)g_1(\phi,t) = 0. \tag{3.16}$$

Let us now deal with h. From Lemma 2, we have

$$h(\phi,t) = \mathbb{E}(\phi(A_t)) = \mathbb{E}(\phi(A_0)) + \mathbb{E}\left(\int_0^t \Big(\phi(1+A_{s-})-\phi(A_{s-})\Big)\lambda_s ds\right). \tag{3.17}$$

From Fubini's theorem, $t \mapsto h(\phi,t)$ is absolutely continuous with respect to Lebesgue's measure and, denoting by $\frac{dh}{dt}$ its derivative,

$$\frac{dh}{dt}(\phi,t) = \mathbb{E}\Big(\big(\phi(1+A_{t-})-\phi(A_{t-})\big)\lambda_t\Big). \tag{3.18}$$

Let $\mathbb{E}^{A_{t-}}$ denote the expectation conditionally to A_{t-}. Then

$$\begin{aligned}\frac{dh}{dt}(\phi,t) &= \mathbb{E}\left(\mathbb{E}^{A_{t-}}\Big(\big(\phi(1+A_{t-})-\phi(A_{t-})\big)\lambda_t\Big)\right)\\ &= \mathbb{E}\Big(\big(\phi(1+A_{t-})-\phi(A_{t-})\big)\mathbb{E}^{A_{t-}}(\lambda_t)\Big).\end{aligned}$$

Now, using the convexity of ϕ, on one hand $\big(\phi(1+A_{t-})-\phi(A_{t-})\big)$ is a nondecreasing function of A_{t-}, and on the other hand, from Assumption 3, $\mathbb{E}^{A_{t-}}(\lambda_t)$ is a nonincreasing function of A_{t-}. Hence $\big(\phi(1+A_{t-})-\phi(A_{t-}), \mathbb{E}^{A_{t-}}(\lambda_t)\big)$ is negatively associated, which ensures that

$$\frac{dh}{dt}(\phi,t) \le \mathbb{E}\big(\phi(1+A_{t-})-\phi(A_{t-})\big)\mathbb{E}(\lambda_t).$$

From the convexity of ϕ

$$\phi(1+A_{t-})-\phi(A_{t-}) \le \phi(1+A_t)-\phi(1+A_t),$$

because $A_{t-} \le A_t$. Therefore

$$\frac{dh}{dt}(\phi,t) \le \mathbb{E}\big(\phi(1+A_t)-\phi(A_t)\big)\mathbb{E}(\lambda_t). \tag{3.19}$$

In other words

$$\frac{dh_0}{dt}(\phi,t) + \big(h_0(\phi,t)-h_1(\phi,t)\big)\mathbb{E}(\lambda_t) \le 0. \tag{3.20}$$

Replacing ϕ by $\phi_a : x \mapsto \phi(x+a)$ in (3.16) and (3.20), we get, for any $a \in \mathbb{R}$,

$$\frac{dg_a}{dt}(\phi,t) + \big(g_a(\phi,t)-g_{a+1}(\phi,t)\big)\mathbb{E}(\lambda_t) = 0, \tag{3.21}$$

$$\frac{dh_a}{dt}(\phi,t) + \big(h_a(\phi,t)-h_{a+1}(\phi,t)\big)\mathbb{E}(\lambda_t) \le 0. \tag{3.22}$$

Now, define, for $u \in \mathbb{R}$ and $x \in \mathbb{R}^+$, the function ϕ_u by

$$\phi_u(x) = (u-x)_+ = \sup(u-x, 0),$$

and consider $\mathcal{E} = \{\phi_u, u \in \mathbb{R}\}$.
It is easy to see that $\phi_u(x+y) = 0$ as soon as $y \geq u$. Hence

$$h_y(\phi_u, t) = g_y(\phi_u, t) = 0 \quad \text{for any } y \geq u. \tag{3.23}$$

Let y be the first integer greater than u. From equation (3.23),

$$h_y(\phi_u, t) \leq g_y(\phi_u, t). \tag{3.24}$$

Now, let us prove, by backward induction on k, that $h_k(\phi_u, t) \leq g_k(\phi_u, t)$ for any k in $[0, y]$.
If $h_k(\phi_u, t) \leq g_k(\phi_u, t)$ at rank k then

$$\frac{dg_{k-1}}{dt}(\phi_u, t) + g_{k-1}(\phi_u, t)\mathbb{E}(\lambda_t) = g_k(\phi_u, t)\mathbb{E}(\lambda_t), \tag{3.25}$$

$$\frac{dh_{k-1}}{dt}(\phi_u, t) + h_{k-1}(\phi_u, t)\mathbb{E}(\lambda_t) \leq g_k(\phi_u, t)\mathbb{E}(\lambda_t). \tag{3.26}$$

Both the initial condition

$$h_{k-1}(\phi_u, 0) = g_{k-1}(\phi_u, 0) = (u-k+1)_+$$

and equations (3.25) and (3.26) imply that $h_{k-1}(\phi_u, t) \leq g_{k-1}(\phi_u, t)$. Hence by induction we get that $h_0(\phi_u, t) \leq g_0(\phi_u, t)$. Whence Theorem 2 is proved for any $\phi \in \mathcal{E}$.
Now, if $\phi \in \mathcal{C}_2$, thanks to Taylor formula, we can write

$$\phi(x) = \phi(0) + x\phi'(0) + \int_0^{+\infty} (x-u)_+ \phi''(u) du. \tag{3.27}$$

Now $(x-u)_+ = (x-u) + (u-x)_+$. Hence equation (3.27) becomes

$$\phi(x) = \phi(0) + x\phi'(0) + \int_0^{\infty} \left((x-u) + (u-x)_+\right) \phi''(u) du. \tag{3.28}$$

Then

$$\mathbb{E}\left(\phi(A_t)\right) = \phi(0) + \mathbb{E}(A_t)\phi'(0) + \mathbb{E}\Big(\int_0^{\infty} \left((A_t-u) + (u-A_t)_+\right) \phi''(u) du\Big).$$

Since the functional inside the integral is nonnegative, Fubini's theorem applies and consequently

$$\mathbb{E}\left(\phi(A_t)\right) = \phi(0) + \mathbb{E}(A_t)\phi'(0) + \int_0^{\infty} \mathbb{E}\Big((A_t-u) + (u-A_t)_+\Big) \phi''(u) du.$$

Now, from the validity of Theorem 2 for the elements of $\mathcal{E}$, we get

$$\mathbb{E}\left(\phi(A_t)\right) \leq \phi(0) + \mathbb{E}(N_t)\phi'(0) + \int_0^{\infty} \mathbb{E}\Big(\left((N_t-u) + (u-N_t)_+\right)\Big) \phi''(u) du.$$

Using again Fubini's theorem we have

$$\mathbb{E}\left(\phi(A_t)\right) \leq \mathbb{E}\left(\phi(N_t)\right). \tag{3.29}$$

We complete the proof using a density argument since $\mathcal{C}_2$ is dense in $\mathcal{C}$. □

3.4 Applications.

In this section, we will give applications of Theorems 1 and 2 of Section 3.2. The first two applications show that suprema of independent binomial random variables (resp. Poisson variables) satisfy a binomial (resp. Poissonian) convex concentration inequality. The third deals with 3−ary search trees. We will show that the process of saturated nodes in an 3−ary search tree is an easy example of discrete time models which satisfied Assumption 1.

3.4.1 Supremum of binomial random variables

Let $p_1 \geq p_2 \geq \ldots \geq p_l$ be a nonincreasing sequence of reals. The aim of this section is to give a concentration inequality for $Z = \sup(B_1, \ldots, B_l)$ where the B_i's are independent $B(n, p_i)$-distributed random variables.

Theorem 4 . *If $Z = \sup(B_1, \ldots, B_l)$, then, for any convex function ϕ, we have $\mathbb{E}(\phi(Z)) \leq \mathbb{E}(\phi(Y))$, where the B_i's are independent $B(n, p_i)$-distributed random variables and $Y \sim B(nl, \mathbb{E}(Z)/nl)$.*

A discrete time representation

Here, we introduce a discrete time counting process $(Z_u)_{u\in\mathbb{N}}$ such that

$$Z = Z_{nl}. \tag{3.30}$$

Next, we apply Theorem 1 to Z_u with $u = nl$.
Let $X_{ij},\ i = 1..n,\ j = 1..l$ be independent variables such that X_{ij} is Bernoulli distributed with parameter p_j. Assume that the p_j's are nonincreasing. If $u = an + b$ with $0 \leq b < n$, we define Z_u by

$$Z_u = \max\Big(S_n(p_1), S_n(p_2), \ldots, S_n(p_a), S_b(p_{a+1})\Big), \tag{3.31}$$

where $S_m(p_j) = \sum_{i=1}^{m} X_{ij}$ is $B(m, p_j)$−distributed.

Lemma 3 . *$(Z_u)_{u\in\mathbb{N}}$ satisfies the hypothesis of Theorem 1.*

The proof of Lemma 3 requires the two technical lemmas below whose proofs are postponed to the end of the section.

Lemma 4 . *Let Y be a $B(n, p)$−distributed random variable and let us denote by G its distribution function. Then for any $k \geq 1$ we have*

$$G^2(k) - G(k-1)G(k+1) \geq 0. \tag{3.32}$$

Lemma 5 . *Assume that $p_1 \geq p_2$. Set*

$$I_k(p_1, p_2) = \frac{\mathbb{P}\Big(S_j(p_2) < k\Big)\mathbb{P}\Big(S_n(p_1) = k\Big)}{\mathbb{P}\Big(S_j(p_2) = k\Big)\mathbb{P}\Big(S_n(p_1) \leq k\Big)}. \tag{3.33}$$

Then the sequence $(I_k(p_1, p_2))_{k=1,\ldots,j}$ is nondecreasing with respect to k for any $j \in \{0, \ldots, n-1\}$.

Proof of lemma 3.

Define $F(m,j,l) = \mathbb{P}(S_m(p_j) \le k)$,

$$N_{a+1}(k) = P(S_b(p_{a+1}) = k) \prod_{i=1}^{a} F(n,i,k),$$

and, for any $i \in \{1,\dots,a\}$,

$$N_i(k) = F(b,a+1,k-1) \prod_{m=1}^{i-1} F(n,m,k)\mathbb{P}(S_n(p_i) = k) \prod_{m=i+1}^{a} F(n,m,k-1).$$

Set $u_k = \mathbb{P}(Z_{j+1} = k+1 \mid Z_j = k)$. Then

$$u_k = \frac{N_{a+1}(k)}{\sum_{i=1}^{a+1} N_i(k)} \mathbb{P}\big(X_b(p_j = 1)\big). \tag{3.34}$$

Let

$$c_k = 1 + \sum_{i=1}^{a} \frac{N_i(k)}{N_{a+1}(k)}. \tag{3.35}$$

From (3.34), we get that $(u_k)_{k\in\mathbb{N}}$ is nonincreasing if and only if $(c_k)_{k\in\mathbb{N}}$ is nondecreasing. Let

$$v_i(k) = \frac{N_i(k)}{N_{a+1}(k)}.$$

It is enough to prove that each sequence $(v_i(k))_{k\in\mathbb{N}}$ is nondecreasing . From the definition of the numbers $N_1(k),\dots,N_{a+1}(k)$,

$$\begin{aligned} v_i(k) &= \frac{\mathbb{P}(S_j(p_a) < k)}{\mathbb{P}(S_j(p_a) = k)} \frac{\mathbb{P}(S_n(p_i) = k)}{\mathbb{P}(S_n(p_i) \le k)} \frac{\prod_{m=i+1}^{a} \mathbb{P}(S_n(p_m) < k)}{\prod_{m=i+1}^{a} \mathbb{P}(S_n(p_m) \le k)} \\ &= I_k(p_i,p_a) \frac{\prod_{m=i+1}^{a} \mathbb{P}(S_n(p_m) < k)}{\prod_{m=i+1}^{a} \mathbb{P}(S_n(p_m) \le k)}. \end{aligned}$$

Using Lemmas 4 and 5 we get that $(v_i(k))_{k\in\mathbb{N}}$ is a product of two nondecreasing sequences. Consequently $(c_k)_{k\in\mathbb{N}}$ is nondecreasing, which ends up the proof. □

Proofs of the technical lemmas

Proof of lemma 4: It is a well known log-concavity result (see Pinelis [8] for instance). Anyway it can easily be proven by induction that

$$\log G(k) \ge \frac{1}{2}\Big(\log G(k-1) + \log G(k+1)\Big).$$

□

Proof of Lemma 5: We will first prove this lemma when $p_1 = p_2 = p$. Set $j = n-m$, then $I_k(p_1, p_2)$ becomes

$$I_k(p_1,p_2) = \frac{\mathbb{P}\Big(S_{n-m} < k\Big)\mathbb{P}\Big(S_n = k\Big)}{\mathbb{P}\Big(S_{n-m} = k\Big)\mathbb{P}\Big(S_n \le k\Big)}.$$

Set

$$\tilde{I}_k = \frac{(n-m-k)!}{(n-k)!}\frac{\mathbb{P}\Big(S_{n-m} < k\Big)}{\mathbb{P}\Big(S_n \le k\Big)},$$

then

$$I_k(p_1,p_2) = \frac{n!}{(n-m)!}(1-p)^m \tilde{I}_k.$$

Since the factor in front of $\tilde{I}_k$ is independent of k, $(I_k(p_1,p_2))_{k\in\mathbb{N}}$ is nondecreasing if and only if $(\tilde{I}_k)_{k\in\mathbb{N}}$ is nondecreasing. Now let J_k be the inverse of $\tilde{I}_k$. Then $J_k = J_k^{(1)} + J_k^{(2)}$ with

$$\begin{cases} J_k^{(1)} = \dfrac{(n-k)!}{(n-m-k)!}\Big(\dfrac{\mathbb{P}(S_{n-m} < k, S_m \le k - S_{n-m})}{\mathbb{P}(S_{n-m} < k)}\Big), \\ \\ J_k^{(2)} = \dfrac{(n-k)!}{(n-m-k)!}\Big(\dfrac{\mathbb{P}(S_{n-m} = k)\mathbb{P}(S_m = 0)}{\mathbb{P}(S_{n-m} < k)}\Big). \end{cases}$$

Consequently, it is enough to show that $\Big(J_k^{(1)}\Big)_{k\in\mathbb{N}}$ and $\Big(J_k^{(2)}\Big)_{k\in\mathbb{N}}$ are nonincreasing. Set $r_i = \mathbb{P}(S_{n-m} = i)$ and $q_i = \mathbb{P}(S_l \le i)$. Then, setting

$$\gamma_1(k) = \frac{(n-k-1)!}{(n-m-k-1)!}\frac{1}{\sum_{j=0}^{k-1} r_j \sum_{j=0}^{k} r_j},$$

we get

$$\frac{J_k^{(1)} - J_{k+1}^{(1)}}{\gamma_1(k)} = (n-k)\sum_{i=0}^{k-1} r_i q_{k-i} \sum_{j=0}^{k} r_j - (n-m-k)\sum_{i=0}^{k} r_i q_{k-i} \sum_{j=0}^{k-1} r_j.$$

Hence $J_k^{(1)} - J_{k+1}^{(1)}$ has the same sign as

$$\delta_1(k) = \left((n-k)\sum_{i=0}^{k-1} r_i q_{k-i} \sum_{j=0}^{k} r_j - (n-m-k)\sum_{i=0}^{k} r_i q_{k-i} \sum_{j=0}^{k-1} r_j\right).$$

Now

$$\delta_1(k) \ge (n-k) r_k \Big(\sum_{i=0}^{k-1} r_i q_{k-i} - q_0 \sum_{i=0}^{k-1} r_i\Big).$$

The right hand side of this inequality is positive since the sequence $(q_i)_{i\in\mathbb{N}}$ is nondecreasing. Hence $(J_k^{(1)})_{k\in\mathbb{N}}$ is nonincreasing.
Let us deal now with $J_k^{(2)}$. Denoting by F the distribution function of S_{n-m} and setting

$$\gamma_2(k) = \frac{(n-k-1)!}{(n-m-k)!}\frac{(1-p)^m}{F(k-1)F(k)},$$

we have

$$J_k^{(2)} - J_{k+1}^{(2)} = \gamma_2(k)\Big((n-k)\big(F(k)-F(k-1)\big)F(k) \\ - (n-m-k)\big(F(k+1)-F(k)\big)F(k-1)\Big). \tag{3.36}$$

Using Lemma 4, we see that the right hand side of (3.36) is nonnegative. Then $(J_k^{(2)})_{k\in\mathbb{N}}$ is nonincreasing, whence $(J_k)_{k\in\mathbb{N}}$ is nonincreasing. Which implies that $(I_k)_{k\in\mathbb{N}}$ is nondecreasing.
Consider now the case where $p_1 > p_2$ and $j = n-m$. Write

$$I_k(p_1,p_2) = I_k(p_1,p_1)\frac{I_k(p_1,p_2)}{I_k(p_1,p_1)}$$

and set

$$L_k := \frac{I_k(p_1,p_2)}{I_k(p_1,p_1)} = \frac{\mathbb{P}(S_j(p_2)<k)\mathbb{P}(S_j(p_1)=k)}{\mathbb{P}(S_j(p_2)=k)\mathbb{P}(S_j(p_1)<k)}.$$

We now prove that $(L_k)_{k\in\mathbb{N}}$ is nondecreasing, which will be enough to conclude.
For $i=1,2$ set $r_i = p_i/(1-p_i)$ (note that since $p_1 > p_2$ we have $r_1 < r_2$). Setting

$$\gamma_3 = \left(\frac{1-p_1}{1-p_2}\right)^j,$$

we get

$$L_k = \gamma_3 \frac{r_1^k}{r_2^k}\frac{\mathbb{P}(S_j(p_2)<k)}{\mathbb{P}(S_j(p_1)<k)}.$$

Expanding L_k, we see that $L_{k+1}-L_k$ as the same sign as Δ_k, with

$$\Delta_k := \frac{\sum_{i=0}^{k-1}\binom{j}{i}r_1^{i+1}}{\sum_{i=0}^{k}\binom{j}{i}r_1^{i}} - \frac{\sum_{i=0}^{k-1}\binom{j}{i}r_2^{i+1}}{\sum_{i=0}^{k}\binom{j}{i}r_2^{i}}.$$

Let

$$C_k(r) = \frac{r\sum_{i=0}^{k-1}\binom{j}{i}r^i}{\sum_{i=0}^{k}\binom{j}{i}r^i} = \frac{rA_{k-1}(r)}{A_{k-1}(r)+\binom{j}{k}r^k}, \tag{3.37}$$

with $A_{k-1}(r) = \sum_{i=0}^{k-1}\binom{j}{i}r^i$. Then it is obvious that $\Delta_k = C_k(r_1) - C_k(r_2)$. Hence Lemma 5 will be proved if we prove that $C_k(r)$ is nondecreasing. Taking the derivative with respect to r in (3.37) we see that the sign of $C_k'(r)$ is the same as the one of

$$d_k(r) = A_{k-1}^2 + \binom{j}{j} r^k\left(rA_{k-1}'(r) - (k-1)A_{k-1}(r)\right).$$

Now

$$d_k(r) = \Big(\sum_{i=0}^{k-1}\binom{j}{i}r^i\Big)^2 - \binom{j}{k}\Big((k-1)+\sum_{i=1}^{k-1}\binom{j}{i}(k-1-i)r^i\Big)r^k \tag{3.38}$$

is a polynomial function in r for which the coefficient of r^{k+i} is

$$\sum_{u=i+1}^{k-1} \binom{j}{u}\binom{j}{k+i-u} - \binom{j}{k}\binom{j}{i}(k-1-i).$$

For $0 \le i \le 2k-2$ and $i+1 \le u \le k-1$, it is easy to check that

$$\binom{j}{i+1}\binom{j}{k-1} \le \binom{j}{u}\binom{j}{k+i-u},$$

whence

$$\binom{j}{k}\binom{j}{i} \le \binom{j}{u}\binom{j}{k+i-u}. \tag{3.39}$$

This last inequality implies that $d_k(r)$ is nonnegative. Which concludes the proof of Lemma 5. □

3.4.2 Supremum of Poisson random variables

Let $\mu_1 \ge \ldots \ge \mu_p$ be a finite nonincreasing sequence of real numbers. The aim of this section is to give a concentration inequality for

$$W = \sup(Y_1, \ldots, Y_p)$$

where the Y_i's are independent and $\mathcal{P}(\mu_i)-$distributed.

Theorem 5 . *For any convex function ϕ we have if $Y \sim \mathcal{P}(\mathbb{E}(W))$ the following inequality*

$$\mathbb{E}(\phi(W)) \le \mathbb{E}(\phi(Y)). \tag{3.40}$$

A continuous time model representation

Here, we introduce a continuous time counting process $(A_t)_{t\ge0}$, such that

$$W = A_1. \tag{3.41}$$

Next, we will apply Theorem 2 to A_t with $t = 1$.
Let μ be equal to $\mu_1 + \ldots + \mu_p$. Let $(T_i)_{i\in\mathbb{N}^*}$ be i.i.d random variables $\mathcal{E}(\mu)-$distributed. Define the process S_n by $S_0 = 0$ and for any $n > 0$, $S_n = \sum_{j=1}^{n} T_j$. It is well known that $N_t = \sum_{k=1}^{+\infty} \mathbb{1}_{\{S_k \le t\}}$ is a Poisson point process. Consider now the nonincreasing sequence of real numbers $(t_i)_{1\le i\le p}$ with sum 1 defined by $t_i\mu = \mu_i$. Let $a_i = \sum_{j=1}^{i} t_i$. Then $N_{a_i} - N_{a_{i-1}}$ is $\mathcal{P}(\mu_i)-$distributed. By homogeneity we can assume that $\mu = 1$ in the sequel. We define for $t \le 1$,

$$N^{(i)} = N_{a_i} - N_{a_{i-1}} \tag{3.42}$$

and

$$k(t) = \sup(i,\ a_i \le t).$$

We consider

$$A_t = \sup(N^{(1)}, \ldots, N^{(k(t))}, N_t - N_{a_{k(t)}}).$$

Lemma 6 *(Compensator of A_t). If $a_i \le t < a_{i+1}$ define λ_t by setting $\lambda_t = 1$ if $A_t = N_t - N_{a_{k(t)}}$ and $\lambda_t = 0$ otherwise. Then $\Lambda_t = \int_0^t \lambda_u du$ is the compensator of A_t.*

Proof of Lemma 6: Let t be in $[a_i, a_{i+1}[$ and $s < t$ we will show that

$$\mathbb{E}^{\mathcal{F}_s}\left(A_t - A_s - \int_s^t \lambda_u du\right) = 0. \tag{3.43}$$

Suppose first that $s > a_i$. Then

$$A_t = \sup\{A_{a_i}, N_t - N_{a_i}\}, \tag{3.44}$$
$$A_s = \sup\{A_{a_i}, N_s - N_{a_i}\}. \tag{3.45}$$

If $A_s = N_s - N_{a_i}$, we have $A_t = N_t - N_{a_i}$ and $\lambda_u = 1$ for any $u \in [s, t]$. As the event

$$B := \{A_s = N_s - N_{a_i}\} \tag{3.46}$$

is $\mathcal{F}_s$-measurable and $N_t - N_s$ is independent of $\mathcal{F}_s$, we have

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_s}\left(\left(A_t - A_s - \int_s^t \lambda_u du\right)\mathbb{1}_{\{B\}}\right) &= \mathbb{E}^{\mathcal{F}_s}\left(\left(N_t - N_s - (t-s)\right)\mathbb{1}_{\{B\}}\right)\\ &= \mathbb{1}_{\{B\}}\mathbb{E}^{\mathcal{F}_s}\left(N_t - N_s - (t-s)\right)\\ &= \mathbb{1}_{\{B\}}\mathbb{E}\left(N_t - N_s - (t-s)\right) = 0.\end{aligned} \tag{3.47}$$

Now on B^c we have $N_s - N_{a_i} < A_s$, whence $A_s = A_{a_i}$. Then

$$\delta := A_{a_i} - (N_s - N_{a_i})$$

is a positive number and is $\mathcal{F}_s$-measurable. Now if $N_t - N_s < \delta$ we have $A_t = A_{a_i}$ and $\lambda_u = 0$ for all $u \in [s, t]$. This implies

$$\mathbb{E}^{\mathcal{F}_s}\left(\left(A_t - A_s - \int_s^t \lambda_u du\right)\mathbb{1}_{\{B^c\}}\mathbb{1}_{\{N_t - N_s < \delta\}}\right) = 0. \tag{3.48}$$

If $N_t - N_s \ge \delta$ let τ be the first time in $]s, t]$ such that $N_\tau - N_{a_i} = A_{a_i}$. Note that τ is a $(\mathcal{F}_t)_{t\ge 0}$-stopping time. Then $A_t = N_t - N_{a_i}$, $A_s = A_{a_i}$, $\lambda_u = 0$ if $u \in [s, \tau[$ and $\lambda_u = 1$ if $u \in [\tau, t]$. Hence

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_s}\Big(\big(A_t - A_s - \int_s^t \lambda_u du\big)\mathbb{1}_{\{B^c\}}\mathbb{1}_{\{N_t - N_s \ge \delta\}}\Big) &=\\ \mathbb{E}^{\mathcal{F}_s}\left(\left(N_t - N_\tau - (t-\tau)\right)\mathbb{1}_{\{B^c\}}\mathbb{1}_{\{N_t - N_s \ge \delta\}}\right).\end{aligned} \tag{3.49}$$

Clearly $\{N_t - N_s > \delta\} = \{N_t - N_\tau > 0\} = \{\tau \le t\}$. Therefore

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_s}\Big(\big(A_t - A_s - \int_s^t \lambda_u du\big)\mathbb{1}_{\{B^c\}}\mathbb{1}_{\{N_t - N_s \ge \delta\}}\Big) &=\\ \mathbb{1}_{\{B^c\}}\mathbb{E}^{\mathcal{F}_s}\left(\mathbb{E}^{\mathcal{F}_\tau}\Big(\big(N_t - N_\tau - (t-\tau)\big)\mathbb{1}_{\{N_t - N_\tau \ge 0\}}\Big)\right) = 0.\end{aligned} \tag{3.50}$$

Putting together equations (3.47) (3.48) (3.50), we get

$$\mathbb{E}^{\mathcal{F}_s}\left(A_t - A_s - \int_s^t \lambda_u du\right) = 0, \tag{3.51}$$

for any $a_i < s < t < a_{i+1}$. Using similar arguments we see that equation (3.51) is valid for any $0 < s < t < 1$. □

Thanks to the following lemma, Theorem 5 is an easy consequence of Theorem 2, with $t = 1$.

Lemma 7 . $(A_t, \lambda_t)_t$ *satisfies Assumption 3*

Proof of Lemma 7: First, for a fixed t, $A_t = A_{t-}$ almost surely. Hence $\mathbb{E}(\lambda_t \mid A_{t-}) = \mathbb{E}(\lambda_t \mid A_t)$ almost surely. If $t \in [a_i,\ a_{i+1}[$, then $\lambda_t = 1$ iff $A_t = N_t - N_{a_i}$. Consequently

$$\mathbb{E}(\lambda_t \mid A_t) = \mathbb{P}(A_t = N_t - N_{a_i} \mid A_t). \tag{3.52}$$

We now prove that the sequence $(u_k(t))_k$ defined below is nonincreasing. Let

$$u_k(t) := \mathbb{P}(A_t = N_t - N_{a_i} \mid A_t = k). \tag{3.53}$$

If $V_i = N_t - N_{a_i}$ then

$$u_k(t) = \frac{\mathbb{P}(V_i = k, N^{(j)} \leq k,\ j = 1..i)}{\mathbb{P}(A_t = k)}. \tag{3.54}$$

The end of the proof needs the following lemma whose proof is postponed to the end of the section.

Lemma 8 . (i) *For any* $1 \leq j \leq i$,

$$v_k(j) = \frac{\mathbb{P}(V_i = k,\ N^{(j)} \leq k)}{\mathbb{P}(V_i = k,\ N^{(j)} \leq k) + \mathbb{P}(V_i < k,\ N^{(j)} = k)}, \tag{3.55}$$

is nonincreasing with respect to k.
(ii) *For any* $i > 0$,

$$M_i(k) = \frac{\mathbb{P}(N^{(i)} < k)}{\mathbb{P}(N^{(i)} \leq k)} \tag{3.56}$$

is nondecreasing with respect to k.

Recall we have to show that $(u_k(t))_{k\in\mathbb{N}}$ is nonincreasing. From the independence of the random variables $N^{(j)}$ and $N_t - N_{a_i}$, setting

$$\begin{aligned} D(k) =& \mathbb{P}(V = k)\prod_{j=1}^{i}\mathbb{P}(N^{(j)} \leq k) \\ &+ \mathbb{P}(V < k)\sum_{u=0}^{i-1}\mathbb{P}(N^{(i-u)} = k)\prod_{l>i-u}\mathbb{P}(N^{(l)} < k)\prod_{l<i-u}\mathbb{P}(N^{(l)} \leq k), \end{aligned}$$

we get

$$u_k(t) = \frac{\mathbb{P}(V = k) \prod_{j=1}^{i} \mathbb{P}(N^{(j)} \leq k)}{D(k)}. \tag{3.57}$$

Set $W_k(t) = (u_k(t))^{-1}$. Rewriting (3.57) we have

$$\begin{aligned} W_k =& 1 + \frac{\mathbb{P}(V < k)\mathbb{P}(N^{(i)} = k)}{\mathbb{P}(V = k)\mathbb{P}(N^{(i)} \leq k)} + \frac{\mathbb{P}(V < k)\mathbb{P}(N^{(i)} < k)\mathbb{P}(N^{(i-1)} = k)}{\mathbb{P}(V = k)\mathbb{P}(N^{(i)} \leq k)\mathbb{P}(N^{(i-1)} \leq k)} \\ &+ \cdots + \frac{\mathbb{P}(V < k)\mathbb{P}(N^{(i)} < k) \ldots \mathbb{P}(N^{(2)} < k)\mathbb{P}(N^{(1)} = k)}{\mathbb{P}(V = k)\mathbb{P}(N^{(i)} \leq k) \ldots \mathbb{P}(N^{(1)} \leq k)}. \end{aligned}$$

Any component of this sum is of the form $\alpha_k \beta_k$, with

$$\alpha_k = \frac{\mathbb{P}(V < k)\mathbb{P}(N^{(j)} = k)}{\mathbb{P}(V = k)\mathbb{P}(N^{(j)} \leq k)} \text{ and } \beta_k = \prod_l \frac{\mathbb{P}(N^{(l)} < k)}{\mathbb{P}(N^{(l)} \leq k)}.$$

Using Lemma 8, we see that both α_k and β_k are nondecreasing sequences of k, which ends the proof of Lemma 7. □

Proof of lemma 8

(i) Let $t \in [a_i,\ a_{i+1}[$. Let us prove that $(v_k(j))_k$ is nonincreasing for all $j \leq i$. Recall that

$$v_k(j) = \frac{\mathbb{P}(V_i = k,\ N^{(j)} \leq k)}{\mathbb{P}(V_i = k,\ N^{(j)} \leq k) + \mathbb{P}(V_i < k,\ N^{(j)} = k)}. \tag{3.58}$$

Fix j and set $v_k = v_k(j)$. It is enough to show that

$$U_k = \frac{\mathbb{P}(V < k)\mathbb{P}(N^{(j)} = k)}{\mathbb{P}(V = k)\mathbb{P}(N^{(j)} \leq k)} \tag{3.59}$$

is nondecreasing in k.

As $\theta_{i+1} < t_j$, and $t - a_i < \theta_j$, there exists θ in $]0, 1[$ such that $t - a_i = \theta t_j$. Equation (3.59) becomes now

$$U_k = \frac{\sum_{n=0}^{k-1} \theta^n t_j^n / n!}{\theta^k \sum_{n=0}^{k} t_j^n / n!}. \tag{3.60}$$

Now, we set, $S_j(u) = \sum_{n=0}^{u} t_j^n / n!$ and

$$K(\theta, k) = \theta^{k+1} S_j(k) S_j(k+1).$$

Then

$$(U_{k+1} - U_k) K(\theta, k) = S_j(k) \left(\sum_{n=0}^{k} \frac{\theta^n t_j^n}{n!} \right) - \theta S_j(k+1) \left(\sum_{n=0}^{k-1} \frac{t_j^n \theta^n}{n!} \right).$$

We will show that for $\theta \in]0,1[$ and $t_j > 0$,

$$P(\theta) = S_j(k)\left(\sum_{n=0}^{k}\frac{\theta^n t_j^n}{n!}\right) - \theta S_j(k+1)\left(\sum_{n=0}^{k-1}\frac{t_j^n\theta^n}{n!}\right)$$

is nonnegative. We have

$$P(\theta) = S_j(k) + \sum_{i=1}^{k}\left(\frac{t_j^i}{i!}S_j(k) - \frac{t_j^{i-1}}{(i-1)!}S_j(k+1)\right)\theta^i. \tag{3.61}$$

Let us first simplify the coefficient in front of θ

$$\begin{aligned} P(\theta) &= S_j(k) + \theta\left(S_j(k)t_j - S_j(k+1)\right) \\ &\quad + \sum_{i=2}^{k}\left(\frac{t_j^i}{i!}S_j(k) - \frac{t_j^{i-1}}{(i-1)!}S_j(k+1)\right)\theta^i. \end{aligned} \tag{3.62}$$

Now using the fact that $\theta < 1$ and $\theta > \theta^2$ we get

$$P(\theta) \geq \theta\left(t_j S_j(k-1) + \frac{k t_j^{k+1}}{(k+1)!}\right) + \sum_{i=2}^{k}\left(\frac{t_j^i}{i!}S_j(k) - \frac{t_j^{i-1}}{(i-1)!}S_j(k+1)\right)\theta^i. \tag{3.63}$$

We can do the same thing for each power of θ, and we get that $P(\theta) \geq 0$ for any $\theta \in [0,1]$.
(ii) Let $P_k = \sum_{n=1}^{k} t_i^n/n! = \sum_{n=1}^{k} c_n$, then

$$\begin{aligned} \frac{M_i(k+1)}{M_i(k)} &= \frac{P_k^2}{P_{k+1}P_{k-1}} = \frac{P_{k-1}^2 + c_k(2P_{k-1}+c_k)}{P_{k-1}^2 + c_k(P_{k-1} + c_{k+1}P_{k-1}/c_k)} \\ &= \frac{P_{k-1}^2 + c_k P_{k-1} + c_k P_k}{P_{k-1}^2 + c_k P_{k-1} + c_{k+1}P_{k-1}}. \end{aligned}$$

Expanding the polynomial function

$$\delta_k = c_k P_k - c_{k+1}P_{k-1} = c_k\left(P_k - \frac{t_i}{k+1}P_{k-1}\right),$$

we see that $\delta_k > 0$, which implies that $M_i(k)$ is nondecreasing. □

3.4.3 3-ary search trees

An 3-ary search tree is a data structure that grows by the progressive insertion of keys into a tree with branch factor 3. Each node contains $0, 1, 2$ keys and gives rise to 3 branches as soon as it contains 2 keys. We call saturated a node containing two keys.
For each $i \in \{1,2,3\}$ let $X_n^{(i)}$ denote the number of nodes containing $i-1$ keys after having introduced $n-1$ keys in the tree. The purpose is to give a binomial convex concentration inequality for $X_n^{(3)}$. In other words we have the following theorem.

Theorem 6 . *The number of saturated nodes in an* 3*-ary search tree satisfies a binomial convex concentration inequality, i.e. for any convex function* ϕ*,*

$$\mathbb{E}\left(\phi(X_n^{(3)})\right) \leq \mathbb{E}\left(\phi(Y)\right), \tag{3.64}$$

where Y *is* $B(n, \mathbb{E}(X_n^{(3)})/n)$*-distributed.*

Construction of an $3-$**ary tree**

Let us first recall Chauvin and Pouyanne description of $3-$ary search trees (see [4] for a general description of $m-$ary search trees). One throws a sequence of numbers in $[0,1]$ named keys, uniformly in $[0,1]^{\mathbb{N}^\star}$. The keys are placed one after another in an 3-ary tree (one node root, from each node grow 3 branches). The following recursive rule describes the way a key named k is inserted in the tree.
i) If the root contains strictly less than 2 keys, then k is inserted in the root. One draws usually keys in a root from left to right in increasing order.
ii) If the root is already saturated, *i.e.* if it contains 2 keys named k_1, k_2, ordered such that $k_1 < k_2$, then corresponds to each interval $I_1 =]-\infty, k_1[,\ I_2 =]k_1, k_2[,\ I_3 =]k_2, \infty[$ a subtree being itself an 3-ary search tree. one draws usually the branches corresponding to I_1, I_2, I_3 from left to right. In this situation, k is inserted in the subtree that corresponds to the interval I_j such that $k \in I_j$. Let $\mathcal{F}_n$, the $\sigma-$field generated up to time n. For each $i \in \{1,2,3\}$ and $n \geq 1$, we define $X_n^{(i)}$ as the number of node which contains $i-1$ keys after the insertion of the $n-1-$th key; such nodes are named nodes of type i. Nodes of type 3 are called saturated.

Proof of theorem 6

We are interested in the saturated nodes. We recall the two following equations that can be found in [4].

$$n - 1 = 2X_n^{(3)} + X_n^{(2)}, \tag{3.65}$$

$$n = X_n^{(1)} + 2X_n^{(2)}. \tag{3.66}$$

Hence if $X_n^{(3)}$ is known then $X_n^{(1)}$ and $X_n^{(2)}$ are also known. It is clear that $X_{n+1}^{(3)} = X_n^{(3)}$ or $X_{n+1}^{(3)} = X_n^{(3)} + 1$ (the number of saturated nodes is an nondecreasing function). $X_n^{(3)}$ increases only if the n-th keys is added in a node of type 2, this is done with probability $\frac{2}{n}X_n^{(2)}$. Hence

$$\begin{aligned} \mathbb{P}(X_{n+1}^{(3)} = X_n^{(3)} + 1 \mid X_n^{(3)}) &= \mathbb{P}(X_{n+1}^{(3)} = X_n^{(3)} + 1 \mid X_n^{(1)}, X_n^{(2)}) \\ &= \frac{2}{n}X_n^{(2)} = \frac{2}{n}(n - 1 - 2X_n^{(3)}). \end{aligned}$$

The last equation implies that $(X_n^{(3)})_{n\in\mathbb{N}}$ satisfies Assumption 1. Theorem 6 follows. □

Remark 2 . *We proved that for* $m = 3$*, the process of saturated nodes* $\left(X_n^{(m)}\right)_{n\in\mathbb{N}}$ *satisfies a binomial convex concentration inequality. The problem to know if this concentration inequality holds for* $m > 3$ *is open.*

Bibliography

[1] V. Bentkus. *on Hoeffding's inequality* To appear in Ann. of Probab. 2003

[2] P. Brémaud. Point processes and queues. Springer-Verlag. 1981.

[3] J. Bretagnolle. Statistique de Kolmogorov-Smirnov pour un échantillon non équiréparti. Dans aspects statistiques et aspects physiques des processus gaussiens. *Paris: Editions du centre nationnal de la recherche scientifique.* 1981.

[4] B. Chauvin, N. Pouyanne. $m-$ary search trees when $m > 26$: a strong asymptotics for the space requirements. Submitted to *Random Structures and Algorithms* 2002.

[5] C. Dellacherie, P.A. Meyer. Probabilités et potentiel. Chap. 5 to 8. Hermann, Paris. 1980.

[6] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc* 58, 13-30. 1963.

[7] I. Pinelis. Optimal Bounds for the Distributions of Martingales in Banach Spaces. *Ann. Prob.* Vol. 22. No. 4, 1679-1706 1994

[8] I. Pinelis. Optimal Tail Comparison Based on Comparison Moments *Eberlein, Ernst (ed.) et al., High dimensional probability. Proceedings of the conference, Oberwolfach, Germany, August 1996. Basel: Birkhäuser. Prog. Probab.* Vol. 43, 297-314 1998.

[9] P. Reynaud-Bourret. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theor. relat. Fields.* 126, 103-153 2003.

[10] P. Reynaud-Bourret. Exponential inequalities for counting processes. *preprint of the Georgia Institute of Technology.* 2002.

[11] Q.M. Shao. A Comparison Theorem on Moment Inequalies Between Negatively Associated and Independent Random Variables. *Journ. of Theor. Prob.*, 2000. Vol. 13, 2. 343-356.

[12] G. Shorack, J. Wellner. Empirical processes with applications to statistics. John Wiley and Sons, Inc., New York. 1986.

Chapter 4

Cas des martingales

Convex concentration inequalities for nondecreasing jump processes with bounded jumps

Thierry KLEIN
Université de versailles
45 avenue des Etats-Unis
78035 Versailles cedex
klein@math.uvsq.fr

Abstract 1 *In this paper, we prove convex concentration inequalities for continuous time nondecreasing jump processes with bounded jumps. These results can be seen as the continuous time analogues of Bentkus result in [1].*

AMS Classification: 60E15, 60F10.
Keywords: Martingales, counting processes, records, convex concentration inequality.

4.1 Introduction.

In this paper, we give convex concentration inequalities for the martingales associated to continuous time nondecreasing jump processes with bounded jumps.

Definition 1 *We say that a random variable X, is more concentrated for the convex functions than a random variable Y if for any convex function ϕ*

$$\mathbb{E}\Big(\phi(X)\Big) \leq \mathbb{E}\Big(\phi(Y)\Big).$$

This concept was first introduced by Hoeffding in [7]. In this paper, Hoeffding compares $\mathbb{E}\left(\phi(S_n)\right)$ with $\mathbb{E}\left(\phi(S_n^\star)\right)$, when $S_n = \sum_{i=1}^n X_i$ is the sum of independent Bernoulli distributed random variables with parameters p_i and $S_n^\star$ is $B(n, \bar{p})-$distributed with $\bar{p}$ the arithmetic mean of the p_i's.

Proposition 1 *(Hoeffding [7], Shorack-Wellner [17]). Let $b_1, \ldots, b_n$ be independent random variables Bernoulli distributed with parameters p_i and $S_n = b_1 + \cdots + b_n$. Let $\bar{p} = \dfrac{p_1 + \ldots + p_n}{n}$ then for any convex function ϕ we have*

$$\mathbb{E}\Big(\phi(S_n)\Big) \leq \mathbb{E}\Big(\phi\big(B(n, \bar{p})\big)\Big). \tag{4.1}$$

These inequalities are very useful to derive tail inequalities as pointed by Hoeffding [7], Bretagnolle [3] who gave a functional version of this result, Pinelis in [11] and [12] studies a more general case where the function ϕ is in a general class of functions. Shao in [16] treats the case of Negatively Associated (N.A) random variables and shows how convex concentrations inequalities lead to classical inequalities like Rosenthal maximal inequality or Kolmogorov inequality. In particular, he is able to extend Hoeffding's inequality on the probability bounds for the sum of a random sample without replacement from a finite population. Bentkus in [1] uses convex concentration inequalities to give bounds for probabilities tails of discrete martingales with bounded jumps. Klein in [9] give convex concentration inequalities for discrete and continuous time counting processes.

In this paper, we give the analogue of Bentkus result when looking at continuous time martingales. These results generalize the results in [9] since the jumps of the processes considered here are bounded by 1. Let $(A_t)_{t\geq 0}$ be a pure bounded jump increasing process i.e. $(A_t)_{t\geq 0}$ is piecewise constant and $0 \leq A_t - A_{t-} \leq 1$ (see Dellacherie-Meyer [6] or Brémaud [2] for complete study of these processes and in particular for properties of their compensator). We note $(C_t)_{t\geq 0}$ is predictable compensator[1]. Then we prove (see Theorem 1) that, for any convex function ϕ

$$\mathbb{E}\left(\phi(A_t - C_t)\right) \leq \mathbb{E}\left(\phi(N_C - C)\right), \tag{4.2}$$

where N_C is $\mathcal{P}(C)-$distributed and $C \geq \|C_t\|_\infty$. The variance of the majorizing Poisson random variable $N_C - C$ is related to the compensator of A_t, and, consequently, may be much greater than the optimal variance factor. Consequently, in Section 4.3, we give concentration inequalities involving the quadratic compensator of $(A_t - C_t)$. However, in that case, the comparison inequality holds only for convex functions with nondecreasing second derivative. Consequently, one can derive only right hand side deviations inequalities from this inequality. This is the reason why we give in Section 4.4 another comparison inequality involving convex function with nonincreasing second derivative, in which, we compare $(A_t - C_t)$ with gaussian random variables. As pointed above convex concentration inequalities are very usefull to derive left and right more precise concentration inequalities as in Pinelis [12] and in Bentkus [1] in the discrete case. For continuous time processes and exponential inequalities derived from bands on the Laplace transform, we refer to Liptser and Shiryaev [10], Courbot [4, 5] and to Wu [18], Houdré and Privault in [8] and Reynaud-Bourret [13, 14] who obtained via the log-Sobolev method concentration inequalities for Poisson processes.

4.2 Expectation type result.

Let $(B_s, \lambda_s)_{s\geq 0}$ be a point process, where the intensity λ_s satisfies $0 \leq \lambda_s \leq 1$. Consider a measurable process $(J_s)_{s\geq 0}$ such that $0 \leq J_s \leq 1$, and N be a poisson point process independent of the other processes.

Definition 2 *We call pure jump process, a process $(A_s)_{s\geq 0}$ of the form*

$$A_s = \int_0^s J_{u-} dB_u.$$

[1] Recall that $A_s - C_s$ is a $\mathcal{F}_s$-martingale.

Remark 1 *The process* $(A_s)_{s\geq 0}$ *is piecewise constant, jumps when the process* B *jumps with jumps equal to* J_{s-}. *The predictable compensator of* A_s *is* $C_t = \int_0^t \lambda_s J_{s-} ds$.

Assumption 1 *Assume that there is some constant* $C > 0$ *such that* $\|C_t\|_\infty \leq C$.

Theorem 1 *Let* $(A_s)_{s\geq 0}$ *be a pure jump process, then under Assumption 1, for any convex function* ϕ, *such that* $\lim_{x\to+\infty}\big(\phi(x)\exp(-Ax)\big) = 0$ *for some positive* A:
a) The function

$$g(s) = \mathbb{E}\Big(\phi\big(A_s - N_{C_s} + N_C - C\big)\Big) \tag{4.3}$$

is nonincreasing.
b) Consequently

$$\mathbb{E}\Big(\phi(A_t - C_t)\Big) \leq \mathbb{E}\Big(\phi(N_C - C)\Big). \tag{4.4}$$

Proof of Theorem 1: Throughout the proof we assume that $\lim_\infty x^{-1}\phi(x) < +\infty$. The general case follows via the Beppo Levi lemma.
Proof of a) Conditionning on $\mathcal{F}_s$, we get

$$g(s) = \mathbb{E}\Big(\mathbb{E}^{\mathcal{F}_s}\big(\phi(N_C - N_{C_s} + A_s - C)\big)\Big). \tag{4.5}$$

Since $N_C - N_{C_s}$ is independent of $\mathcal{F}_s$ and is $\mathcal{P}(C - C_s)-$distributed,

$$g(s) = \mathbb{E}\Big(\sum_{k=0}^{\infty}\exp(C_s - C)\frac{(C - C_s)^k}{k!}\phi(k + A_s - C)\Big). \tag{4.6}$$

Applying Fubini's Theorem we have

$$g(s) = \sum_{k=0}^{\infty}\mathbb{E}\Big(\exp(C_s - C)\frac{(C - C_s)^k}{k!}\phi(k + A_s - C)\Big). \tag{4.7}$$

For $k \geq 0$, let

$$\begin{cases} f_k(s) = \exp(C_s - C)\dfrac{(C - C_s)^k}{k!}, \\ \\ h_k(s) = \phi(k + A_s - C). \end{cases}$$

Since the function f_k is a.c. and h_k is of bounded variation we may apply the product formula for Stieltjes integral (see Dellacherie Meyer [6] or Brémaud [2]) i.e.

$$f_k(s)h_k(s) - f_k(0)h_k(0) = \int_0^s \Big(f_k(u)dh_k(u) + h_k(u-)df_k(u)\Big). \tag{4.8}$$

Note that $\mathbb{E}\sum_{k=0}^\infty f_k(0)h_k(0) = g(0)$, hence,

$$g(s) = g(0) + \sum_{k=0}^{\infty}\mathbb{E}\int_0^s f_k(u)dh_k(u) + \sum_{k=0}^{\infty}\mathbb{E}\int_0^s h_k(u-)df_k(u). \tag{4.9}$$

In order to have more readable equations, we set $X^{(k)}_{u-} = A_{u-} + k - C$, then

$$h_k(s) = h_k(0) + \int_0^s \Big(\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})\Big)\frac{dA_u}{J_{u-}}, \tag{4.10}$$

and

$$\begin{cases} f_0(s) = \displaystyle\int_0^s \exp(C_u - C)\lambda_u J_{u-} du, \\ f_k(s) = \displaystyle\int_0^s \exp(C_u - C)\Big(\frac{(C-C_u)^k}{k!} - \frac{(C-C_u)^{k-1}}{(k-1)!}\Big)\lambda_u J_{u-} du \quad \text{for} \quad k \geq 1. \end{cases}$$

Consequently,

$$\int_0^s f_k(u) dh_k(u) = \int_0^s f_k(u)\Big(\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})\Big)\frac{dA_u}{J_{u-}}, \tag{4.11}$$

$$\int_0^s h_0(u-) df_0(u) = \int_0^s h_0(u-)\exp(C_u - C)\lambda_u J_{u-} du, \tag{4.12}$$

$$\int_0^s h_k(u-) df_k(u) = \int_0^s h_k(u-)\exp(C_u - C)\Big(\frac{(C-C_u)^k}{k!} - \frac{(C-C_u)^{k-1}}{(k-1)!}\Big)\lambda_u J_{u-} du \text{ for } k \geq 1. \tag{4.13}$$

Since $X^{(k+1)}_{u-} = X^{(k)}_{u-} + 1$, equations (4.12) and (4.13) can be written in the following way

$$\int_0^s h_k(u-) df_k(u) = \int_0^s \big(h_k(u-) - h_{k+1}(u-)\big)\exp(C_u - C)\frac{(C-C_u)^k}{k!}\lambda_u J_{u-} du.$$

From Theorem T8 of Brémaud page 27 in [2] and the fact that

$$f_k(u)\Big(\frac{\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})}{J_{u-}}\Big)$$

is predictable, the compensator of the process

$$\int_0^s f_k(u)\Big(\frac{\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})}{J_{u-}}\Big) dA_u$$

is equal to

$$\int_0^s f_k(u)\Big(\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})\Big)\lambda_u du,$$

which ensures that

$$\mathbb{E}\int_0^s f_k(u) dh_k(u) = \mathbb{E}\int_0^s f_k(u)\Big(\phi(J_{u-} + X^{(k)}_{u-}) - \phi(X^{(k)}_{u-})\Big)\lambda_u du. \tag{4.14}$$

Hence

$$g(s) = g(0) + \sum_{k=0}^{\infty} \mathbb{E} \int_0^s f_k(u)\Big(\big(\phi(J_{u-} + X_{u-}^{(k)}) - \phi(X_{u-}^{(k)})\big) - J_{u-}\big(\phi(1 + X_{u-}^{(k)}) - \phi(X_{u-}^{(k)})\big)\Big)\lambda_u du. \tag{4.15}$$

Using the convexity of ϕ we see that

$$\phi(J_{u-} + X_{u-}^{(k)}) - \phi(X_{u-}^{(k)}) \leq J_{u-}\Big(\phi(1 + X_{u-}^{(k)}) - \phi(X_{u-}^{(k)})\Big).$$

It follows that $g(s)$ is nonincreasing. hence a) is proved.
Proof of b): Using a) we get for any $t \geq 0$

$$\mathbb{E}\Big(\phi\big(A_t - N_{C_t} + N_C - C\big)\Big) \leq \mathbb{E}\Big(\phi\big(N_C - C\big)\Big). \tag{4.16}$$

Now we apply Jensen inequality for the conditional expectation with respect to $\mathcal{F}_t$:

$$\mathbb{E}\phi\Big(\mathbb{E}^{\mathcal{F}_t}\big(A_t - N_{C_t} + N_C - C\big)\Big) \leq \mathbb{E}\Big(\mathbb{E}^{\mathcal{F}_t}\phi\big(A_t - N_{C_t} + N_C - C\big)\Big). \tag{4.17}$$

Since A_t is $\mathcal{F}_t$-measurable and $N_C - N_{C_t}$ is independent of $\mathcal{F}_t$, the left hand side of equation (4.17) is equal to $\mathbb{E}\big(\phi(A_t + C_t)\big)$. This ends the proof of b). $\square$

4.3 Variance type result.

We introduce the quadratic process

$$D_s = \int_0^s \lambda_u J_{u-}^2 du, \tag{4.18}$$

and we make the following assumption.

Assumption 2 . *Assume that there is some constant $D > 0$ such that $\|D_t\|_\infty \leq D$.*

Theorem 2 . *Let $(A_s)_{s\geq 0}$ be a pure jump process, then under Assumption 2, for any C^2 convex function ϕ with nondecreasing second derivative, such that $\lim_{+\infty} \phi(x)e^{-Ax} = 0$ for some positive A.*
a) The function

$$h(s) = \mathbb{E}\Big(\phi(A_s - N_{D_s} + N_D - D + D_s - C_s)\Big). \tag{4.19}$$

is nonincreasing.
b) Consequently,

$$\mathbb{E}\Big(\phi(A_t - C_t)\Big) \leq \mathbb{E}\Big(\phi(N_D - D)\Big). \tag{4.20}$$

Remark 2 *Theorem 2 recovers right-hand side concentration inequalities. For instance, following Bentkus proof in [1], on can prove*

$$\mathbb{P}\Big(A_t - C_t \geq x\Big) \leq \frac{e^2}{2}\mathbb{P}\Big(N_D - D \geq x\Big). \tag{4.21}$$

Note that this inequality is stronger (at least for large values of x) than the following classical exponential inequality,

$$\mathbb{P}\Big(A_t - C_t \geq x\Big) \leq \exp\Big(-Dh(x/D)\Big), \tag{4.22}$$

where $h(u) = (1+u)\log(1+u) - u$.

Proof of Theorem 2: The proof is similar to the proof of Theorem 1. Throughout the proof we assume that $\lim_{+\infty} x^{-1}\phi(x) < +\infty$. The general case follows via the Beppo Levi lemma. Take first the conditional expectation with respect to $\mathcal{F}_s$

$$g(s) = \mathbb{E}\Big(\exp(D_s - D)\sum_{k=0}^{\infty}\frac{(D-D_s)^k}{k!}\phi(A_s + D_s - C_s - D + k)\Big). \tag{4.23}$$

Define now

$$\begin{cases} f_k(s) = \exp(D_s - D)\dfrac{(D-D_s)^k}{k!}, \\ \\ h_k(s) = \phi(A_s - C_s + D_s - D + k). \end{cases}$$

Using again the product formula and since $\mathbb{E}\sum_{k=0}^{\infty} f_k(0)h_k(0) = g(0)$, we have

$$g(s) = g(0) + \sum_{k=0}^{\infty}\mathbb{E}\int_0^s f_k(u)dh_k(u) + \sum_{k=0}^{\infty}\mathbb{E}\int_0^s h_k(u-)df_k(u). \tag{4.24}$$

Let $Y_{u-}^{(k)} = A_{u-} + D_u - C_u + k - D$, as in the previous section we have

$$h_k(s) = \int_0^s \Big(\frac{\phi(J_{u-} + Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)})}{J_{u-}}\Big)dA_u - \int_0^s \phi'(Y_{u-}^{(k)})(1 - J_{u-})\lambda_u J_{u-}du, \tag{4.25}$$

$$f_0(s) = \int_0^s \exp(D_u - D)\lambda_u J_{u-}^2 du, \tag{4.26}$$

$$f_k(s) = \int_0^s \exp(D_u - D)\Big(\frac{(D-D_u)^k}{k!} - \frac{(D-D_u)^{k-1}}{(k-1)!}\Big)\lambda_u J_{u-}^2 du \text{ for } k \geq 1. \tag{4.27}$$

Consequently

$$\int_0^s f_k(u)dh_k(u) = \int_0^s f_k(u)\frac{\phi(J_{u-} + Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)})}{J_{u-}}dA_u - \int_0^s f_k(u)\phi'(Y_{u-}^{(k)})(1 - J_{u-})\lambda_u J_{u-}du, \tag{4.28}$$

$$\int_0^s h_0(u-)df_0(u) = \int_0^s h_0(u-)\exp(D_u - D)\lambda_u J_{u-}^2 du, \text{ and for } k \geq 1, \tag{4.29}$$

$$\int_0^s h_k(u-)df_k(u) = \int_0^s h_k(u-)\exp(D_u - D)\Big(\frac{(D-D_u)^k}{k!} - \frac{(D-D_u)^{k-1}}{(k-1)!}\Big)\lambda_u J_{u-}^2 du. \tag{4.30}$$

Since $Y_{u-}^{(k+1)} = Y_{u-}^{(k)} + 1$, the two last equations can be written in the following way

$$\int_0^s h_k(u-)df_k(u) = \int_0^s \big(h_k(u-) - h_{k+1}(u-)\big)\exp(D_u - D)\frac{(D-D_u)^k}{k!}\lambda_u J_{u-}^2 du. \quad (4.31)$$

From Theorem T8 of Brémaud page 27 in [2] and the fact that

$$f_k(u)\frac{\phi(J_{u-} + Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)})}{J_{u-}}$$

is predictable, the compensator of the process

$$\int_0^s f_k(u)\frac{\phi(J_{u-} + Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)})}{J_{u-}}dA_u$$

is equal to

$$\int_0^s f_k(u)\Big(\phi(J_{u-} + Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)})\Big)\lambda_u du,$$

which ensures that

$$\mathbb{E}\int_0^s f_k(u)dh_k(u) = \mathbb{E}\int_0^s f_k(u)\Big(\phi(J_{u-}+Y_{u-}^{(k)}) - \phi(Y_{u-}^{(k)}) - \phi'(Y_{u-}^{(k)})(1-J_{u-})J_{u-}\Big)\lambda_u du. \quad (4.32)$$

Consider now

$$S(y,a) = \phi(a+y) - \phi(y) - \phi'(y)a(1-a) + \Big(\phi(y) - \phi(1+y)\Big)a^2.$$

If $\bar{g}(s) = g(s) - g(0)$,

$$\bar{g}(s) = \sum_{k=1}^{\infty}\mathbb{E}\int_0^s f_k(u)S(Y_{u-}^{(k)}, J_{u-})\lambda_u du. \quad (4.33)$$

Tthe function inside the integral in equation (4.33) has the same sign as the sign of $S(Y_{u-}^{(k)}, J_{u-})$. Now, rewritting (4.33), we get

$$S(y,a) = \phi(y+a) - \phi(y) - a\phi'(y) - a^2(\phi(y+1) - \phi(y) - \phi'(y)).$$

Using Taylor formula, we have

$$\phi(y+a) - \phi(y) - a\phi'(y) = a^2\int_0^1 (1-t)\phi''(at+y)dt.$$

Since ϕ'' is nondecreasing and $a \in [0,1]$, we have:

$$\phi(y+a) - \phi(y) - a\phi'(y) \leq a^2\int_0^1 (1-t)\phi''(t+y)dt.$$

Since

$$\int_0^1 (1-t)\phi''(t+y)dt = \phi(y+1) - \phi(y) - \phi'(y),$$

it implies that $S(y,a) \leq 0$ for $a \in [0,1]$. Hence g is nonincreasing and a) is proved. The proof of b) is the same as the proof of Theorem 1 b). □

4.4 Brownian Case

Now, it is quite natural to ask if Theorem 2 also provides left-hand-side concentration inequalities for the process $(A_t - C_t)_{t\geq 0}$. Unfortunately the answer is negative (for exanple the second derivative of the function $\phi_t(x) = \exp(tx)$ for nonpositive t is nondecreasing). That is why we introduce a brownian motion in order to get the analogous to equations (4.21) and (4.22) for the left-hand side. Let $(W_s)_{s\geq}$ be a Brownian motion independent of everything else.

Theorem 3 *Let $(A_s)_{s\geq 0}$ be a pure jump process, then under Assumption 2, for any C^2 convex function ϕ with nondecreasing second derivative and such that $\lim_{+\infty} \phi(x)e^{-Ax} = 0$ for some positive A.*
a) The function

$$h(s) = \mathbb{E}\Big(\phi(C_s - A_s + W_D - W_{D_s})\Big) \tag{4.34}$$

is nonincreasing.
b) Consequently,

$$\mathbb{E}\Big(\phi(C_t - A_t)\Big) \leq \mathbb{E}\Big(\phi(W_D)\Big). \tag{4.35}$$

Remark 3 *This Theorem gives bounds for $\mathbb{P}\big(A_t - C_t \leq x\big)$. Indeed,*

$$\mathbb{P}\big(A_t - C_t \leq -x\big) = \mathbb{P}\big(C_t - A_t \geq x\big),$$

set $\phi_\lambda(y) = \exp(\lambda y)$ for any λ positive, ϕ_λ is a convex function with nondecreasing second derivative. We can apply Theorem 3,

$$\begin{aligned}\mathbb{P}\big(A_t - C_t \leq -x\big) &\leq \exp\Big(-\lambda x + L_{W_D}(\lambda)\Big),\\ &\leq \exp\Big(-\lambda x + D\lambda^2/2\Big).\end{aligned}$$

where L_{W_D} is the logarithm of the laplace transform of the variable W_D which is $\mathcal{N}(0, D)$-distributed. Consequently

$$\mathbb{P}\big(A_t - C_t \leq -x\big) \leq \exp\Big(-\frac{x^2}{2D}\Big).$$

Note that Pinelis in [12] derives more precise inequalities from inequalities in the style of (4.35).

Proof: Throughout the proof we assume that $\lim_{+\infty} \phi''(x) < +\infty$. Note that, since ϕ'' is nondeacreasing, $\lim_{-\infty} \phi''(x)$ exixts in $\mathbb{R}_+$. The general case follows via the Beppo Levi lemma.We may assume $D \geq \|D_s\|_\infty + \epsilon$ for some positive ϵ. First we take the conditionnal expectation with respect to $\mathcal{F}_s$

$$h(s) = \mathbb{E}\Big(\int_{\mathbb{R}} \phi(C_s - A_s + x) \exp\Big(-\frac{x^2}{2(D - D_s)}\Big)\frac{dx}{\sqrt{2\pi(D - D_s)}}\Big). \tag{4.36}$$

Applying Fubini Theorem, we get

$$h(s) = \int_{\mathbb{R}} \mathbb{E}\Big(\phi(C_s - A_s + x)\exp\Big(-\frac{x^2}{2(D-D_s)}\Big)\frac{dx}{\sqrt{2\pi(D-D_s)}}\Big). \tag{4.37}$$

Define

$$\begin{cases} f_x(s) = \phi(C_s - A_s + x), \\ g_x(s) = \exp(-\frac{x^2}{2(D-D_s)})\frac{1}{\sqrt{(D-D_s)}}. \end{cases}$$

As previously, we can apply the product formula for Stieltjes integral

$$f_x(s)g_x(s) = f_x(0)g_x(0) + \int_0^s f_x(u-)dg_x(u) + \int_0^s g_x(u)df_x(u). \tag{4.38}$$

Hence

$$h(s) - h(0) = \int_{\mathbb{R}} \mathbb{E}\int_0^s g_x(u)df_x(u)\frac{dx}{\sqrt{2\pi}} + \int_{\mathbb{R}} \mathbb{E}\int_0^s f_x(u-)dg_x(u)\frac{dx}{\sqrt{2\pi}}. \tag{4.39}$$

Set $Z_{u-}^{(x)} = C_u - A_{u-} + x$. As in the previous section we have the following integral representations

$$f_x(s) = \int_0^s \frac{\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)})}{J_{u-}}dA_u + \int_0^s \lambda_u J_{u-}\phi'(Z_{u-}^{(x)})du, \tag{4.40}$$

$$g_x(s) = \frac{1}{2}\int_0^s g_x(u)D'_u\Big(\frac{1}{D-D_u} - \frac{x^2}{(D-D_u)^2}\Big)du. \tag{4.41}$$

Consequently,

$$\int_0^s f_x(u-)dg_x(u) = \frac{1}{2}\int_0^s f_x(u-)g_x(u)D'_u\Big(\frac{1}{D-D_u} - \frac{x^2}{(D-D_u)^2}\Big)du,$$

$$\int_0^s g_x(u)df_x(u) = \int_0^s g_x(u)\frac{\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)})}{J_{u-}}dA_u + \int_0^s g_x(u)\lambda_u J_{u-}\phi'(Z_{u-}^{(x)})du.$$

Let us deal with the first term in the right-hand side of equation (4.39). From Theorem T8 of Brémaud page 27 in [2] and the fact that

$$g_x(u)\frac{\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)})}{J_{u-}}$$

is predictable, the compensator of the process

$$\int_0^s g_x(u)\frac{\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)})}{J_{u-}}dA_u$$

is equal to

$$\int_0^s g_x(u)\Big(\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)})\Big)\lambda_u du,$$

which ensures that

$$\mathbb{E}\int_0^s g_x(u)df_x(u) = \mathbb{E}\int_0^s g_x(u)\Big(\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)}) + J_{u-}\phi'(Z_{u-}^{(x)})\Big)\lambda_u du.$$

So the first term in the right-hand-side of equation (4.39) is equal to

$$\int_{\mathbb{R}} E\int_0^s g_x(u)\Big(\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)}) + J_{u-}\phi'(Z_{u-}^{(x)})\Big)\lambda_u du\frac{dx}{\sqrt{2\pi}}. \tag{4.42}$$

Now we deal with the second term in the right-hand side of equation (4.39). Using the integral representation we get

$$\int_{\mathbb{R}}\mathbb{E}\int_0^s f_x(u-)dg_x(u)\frac{dx}{\sqrt{2\pi}} = \frac{1}{2}\mathbb{E}\int_0^s\int_{\mathbb{R}} f_x(u-)g_x(u)D'_u\Big(\frac{1}{D-D_u} - \frac{x^2}{(D-D_u)^2}\Big)\frac{dx}{\sqrt{2\pi}}du. \tag{4.43}$$

Set

$$I_1(u) = \int_{\mathbb{R}}\phi(Z_{u-}^{(x)})\exp\Big(-\frac{x^2}{2(D-D_u)}\Big)\frac{D'_u}{(D-D_u)^{3/2}}\frac{dx}{\sqrt{2\pi}},$$
$$I_2(u) = \int_{\mathbb{R}}\phi(Z_{u-}^{(x)})x^2\exp\Big(-\frac{x^2}{2(D-D_u)}\Big)\frac{D'_u}{(D-D_u)^{5/2}}\frac{dx}{\sqrt{2\pi}}.$$

Then, the second term in the right-hand side of equation (4.39) is

$$\int_{\mathbb{R}}\mathbb{E}\int_0^s f_x(u-)dg_x(u)\frac{dx}{\sqrt{2\pi}} = \frac{1}{2}\mathbb{E}\int_0^s (I_1(u) - I_2(u))du. \tag{4.44}$$

Next we integrate by parts I_2 twice to get ride of x^2 in the integral. Set

$$\begin{cases} U(x) = x\phi(Z_{u-}^{(x)}), \\ V'(x) = x\exp\Big(-\frac{x^2}{2(D-D_u)}\Big). \end{cases}$$

Since $\frac{\partial Z_{u-}^{(x)}}{\partial x} = 1$ we get

$$I_2(u) = \int_{\mathbb{R}}\Big(x\phi'(Z_{u-}^{(x)}) + \phi(Z_{u-}^{(x)})\Big)\exp\Big(\frac{-x^2}{2(D-D_u)}\Big)\frac{D'_u}{(D-D_u)^{3/2}}\frac{dx}{\sqrt{2\pi}}.$$

Setting

$$I_3(u) = \int_{\mathbb{R}} x\phi'(Z_{u-}^{(x)})\exp\Big(\frac{-x^2}{2(D-D_u)}\Big)\frac{D'_u}{(D-D_u)^{3/2}}\frac{dx}{\sqrt{2\pi}},$$

we have

$$I_2(u) = I_3(u) + I_1(u).$$

Then we integrate I_3 by parts with,

$$\begin{cases} U(x) = \phi(Z_{u-}^{(x)}), \\ V'(x) = x \exp\Big(-\dfrac{x^2}{2(D-D_u)}\Big). \end{cases}$$

We thus get

$$I_3 = \int_{\mathbb{R}} \phi''(Z_{u-}^{(x)}) \exp\Big(\frac{-x^2}{2(D-D_u)}\Big) \frac{D'_u}{\sqrt{D-D_u}} \frac{dx}{\sqrt{2\pi}}. \tag{4.45}$$

Hence

$$\int_{\mathbb{R}} \mathbb{E} \int_0^s f_x(u-) dg_x(u) \frac{dx}{\sqrt{2\pi}} = -\frac{1}{2} \mathbb{E} \int_0^s \int_{\mathbb{R}} \Big(\phi''(Z_{u-}^{(x)}) g_x(u) D'_u \frac{dx}{\sqrt{2\pi}}\Big) du. \tag{4.46}$$

Setting $\bar{h}(t) = h(t) - h(0)$, we get from equations (4.42) and (4.46),

$$\bar{h}(s) = \int_{\mathbb{R}} \mathbb{E} \int_0^s g_x(u) \Big(\phi(Z_{u-}^{(x)} - J_{u-}) - \phi(Z_{u-}^{(x)}) + J_{u-}\phi'(Z_{u-}^{(x)}) - J_{u-}^2 \frac{\phi''(Z_{u-}^{(x)})}{2}\Big) \lambda_u du \frac{dx}{\sqrt{2\pi}}. \tag{4.47}$$

Define

$$l(\delta) = \phi(x-\delta) - \phi(x) + \delta\phi'(x) - \frac{\delta^2}{2}\phi''(x).$$

Then

$$l'(\delta) = \phi'(x) - \phi'(x-\delta) - \delta\phi''(x).$$

Clearly $l(0) = 0$ and $l' \leq 0$ as soon as ϕ' is convex. Consequently, the function l is nonpositive. This conclude the proof of a). The proof of b) is the same as the proof of Theorem 1 b). □

Bibliography

[1] V. Bentkus. *on Hoeffding's inequality* To appear in Ann. of Probab. 2003

[2] P. Brémaud. Point processes and queues. Springer-Verlag. 1981.

[3] J. Bretagnolle. Statistique de Kolmogorov-Smirnov pour un échantillon non équiréparti. Dans aspects statistiques et aspects physiques des processus gaussiens. *Paris: Editions du centre nationnal de la recherche scientifique.* 1981.

[4] B. Courbot. Vitesses de convergence pour les martingales dans le théorème central limite fonctionnel: comparaisons de méthodes, cadres uni et multidimentionnel. Thèse de doctorat del'université de Rennes 1, 1998

[5] B. Courbot. Rates of convergence in the functionnal CLT for martingales. *C.R. Acad. Sci. Paris, t. 328, Série 1, p. 509-513,* 1999.

[6] C. Dellacherie, P.A. Meyer. Probabilités et potentiel, Chap. 5 to 8. Hermann, Paris. 1980.

[7] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc* 58, 13-30. 1963.

[8] C. Houdré, N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli 8, No.6, 697-720.* 2002

[9] T. Klein. Convex concentration inequalities for nondecreasing processes. *Preprint 113 of the LAMA, university of Versailles-Saint-Quentin.* 2003.

[10] R.S. Liptser, A.N. Shiryaev. Theory of martingales. Kluwer Academic Publishers, Dordrecht, 1986.

[11] I. Pinelis. Optimal Bounds for the Distributions of Martingales in Banach Spaces. *Ann. Prob.* Vol. 22. No. 4, 1679-1706 1994.

[12] I. Pinelis. Optimal Tail Comparison Based on Comparison Moments. *Eberlein, Ernst (ed.) et al., High dimensional probability. Proceedings of the conference, Oberwolfach, Germany, August 1996. Basel: Birkhäuser. Prog. Probab.* Vol. 43, 297-314 1998.

[13] P. Reynaud-Bourret. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory. relat. Fields.* 126, 103-153 2003.

[14] P. Reynaud-Bourret. Exponential inequalities for counting processes. *preprint of the Georgia Institute of Technology.* 2002.

[15] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Relat. Fields 119, 163-175*, 2001

[16] Q.M. Shao. A Comparison Theorem on Moment Inequalies Between Negatively Associated and Independent Random Variables. *J. of Theor. Prob.*, 2000. Vol. 13, 2. 343-356.

[17] G. Shorack, J. Wellner. Empirical processes with applications to statistics. John Willey and Sons, Inc., New York. 1986.

[18] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Relat. Fields 119, 427-438*, 2000.

Part IV

Martingales et Arbres aléatoires

Chapter 5

Martingales et Arbres aléatoires

Martingales, Embedding and Tilting of Binary Trees

B. Chauvin , T. Klein , J-F. Marckert , A. Rouault
Université de versailles
45 avenue des Etats-Unis
78035 Versailles cedex
name@math.uvsq.fr

September, the 22th

Abstract 1 *We are interested in the asymptotic analysis of the binary search tree (BST) under the random permutation model. Two methods are mainly used: the first one is the embedding in continuous time and the second one is the tilting probability method. Combining both gives a commutative scheme between four models:*

$$\begin{array}{ccc} \textit{BST} & \xrightarrow{\textit{embedding}} & \textit{Yule tree / fragmentation} \\ \Big\downarrow \textit{tilting} & & \Big\downarrow \textit{tilting} \\ \textit{tilted BST} & \xrightarrow{\textit{embedding}} & \textit{tilted Yule tree / tilted fragmentation} \end{array}$$

In this paper we focus on the upper embedding arrow and on the tilting arrows. We thus get new results on the BST and also new proofs of known results. In particular, thanks to the left tilting arrow, we give a conceptual proof (in the sense of Lyons, Pemantle, Peres) of the asymptotic behavior of the profile.

Key words. Binary search tree, fragmentation, branching random walk, probability tilting, convergence of martingales, Chinese restaurant.

A.M.S. Classification. 60J25, 60J80, 68W40, 60C05, 60G42, 60G44.

5.1 Introduction

5.1.1 The model of binary search trees

For a convenient definition of trees we are going to work with, let first define

$$\mathbb{U} = \maltese \cup \bigcup_{n \geq 1} \{0,1\}^n$$

as the set of finite words on the alphabet $\{0,1\}$ (with $\maltese$ as an empty word). For u and v in $\mathbb{U}$, denote by uv the concatenation of the word u with the word v (by convention we set, for any $u \in \mathbb{U}$, $\maltese u = u$). If $v \neq \maltese$, we say that uv is a descendant of u and u is an ancestor of uv. Moreover $u0$ (resp. $u1$) is called left (resp. right) child of u.

A *complete binary tree* T is a finite subset of $\mathbb{U}$ such that

$$\begin{cases} \maltese \in T \\ \text{ if } uv \in T \text{ then } u \in T\,, \\ u1 \in T \Leftrightarrow u0 \in T\,. \end{cases}$$

The elements of T are called *nodes* , and $\maltese$ is called the *root* ; $|u|$, the number of letters in u, is the *depth* of u (with $|\maltese| = 0$). Write **BinTree** for the set of complete binary trees.

A tree $T \in$ **BinTree** can be described by giving the set ∂T of its *leaves*, that is, the nodes that are in T but with no descendants in T. The nodes of $T \backslash \partial T$ are called *internal* nodes[1].

We study binary search trees (BST), which are widely used to store totally ordered data (the monograph of Mahmoud [28] gives an overview of the state of the art).

Let A be a totally ordered set of elements named keys and for $n \geq 1$, let $(a_1, ..., a_n)$ be picked up without replacement from A. The (labeled) binary search tree built from these data is a complete binary tree in which each internal node is associated with a key belonging to $(a_1, ..., a_n)$ in the following way:

The first key a_1 is assigned to the root. The next key a_2 is assigned to the left child of the root if it is smaller than a_1, or it is assigned to the right child of the root if it is larger than a_1. We proceed further inserting key by key recursively. We get a labeled complete binary tree with n internal nodes such that the keys of the left subtree of any given node u are smaller than the key of u, and the keys of the right subtree are larger than the key of u.

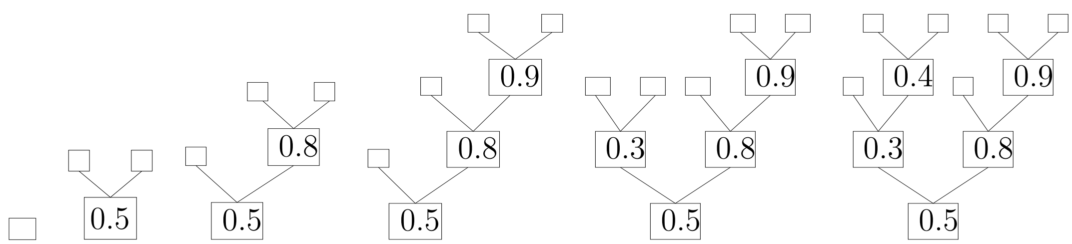

Figure 5.1: BST built with the sequence of data 0.5, 0.8, 0.9, 0.3, 0.4 (empty squares are leaves).

To study the shape of these trees for large n, it is classical to introduce a random model. One usually assumes that the data $(x_i)_{i\geq 1}$ successively inserted are i.i.d. random variables with a continuous distribution F. For every $n \geq 1$, the string $x_1, .., x_n$ induces (a.s.) a permutation σ_n such that $x_{\sigma_n(1)} < x_{\sigma_n(2)} < \cdots < x_{\sigma_n(n)}$. Since the x_i are exchangeable, σ_n is uniformly distributed on the set $\mathcal{S}_n$ of permutations of $\{1, .., n\}$. Since this claim is not sensitive to F we will assume, for the sake of simplicity that F is the uniform distribution on $[0, 1]$. This is the so-called random permutation model.

[1]Some authors consider non complete binary trees, removing the third condition in the above definition. The boundary is then the set of nodes that are not in T but whose predecessors are in T ([6]). It can be seen as a set of external (or available nodes). Here, we choose to work with complete trees, but this choice has no impact on the results.

Again by exchangeability, σ_n is independent of the vector $(x_{\sigma_n(1)}, x_{\sigma_n(2)}, \dots, x_{\sigma_n(n)})$ and we have

$$\begin{aligned} P\big(x_{n+1} \in (x_{\sigma_n}(j), x_{\sigma_n}(j+1)) \,|\, \sigma_n\big) &= P\big(x_{n+1} \in (x_{\sigma_n}(j), x_{\sigma_n}(j+1))\big) \\ &= P(\sigma_{n+1}(j+1) = n+1) = (n+1)^{-1} \end{aligned}$$

for every $j = 0, 1, .., n$, where $x_{\sigma_n(0)} := 0$ and $x_{\sigma_n(n+1)} := 1$. One can also express this property with the help of the sequential ranks of the permutation: the random variables $R_k = \sum_{j=1}^k \mathbb{1}_{x_j \le x_k}, k \ge 1$ are independent and R_k is uniform on $\{1, \dots, k\}$ (see for instance Mahmoud [28], section 2.3) , so that $P(R_{n+1} = j+1 \,|\, R_1, .., R_n) = (n+1)^{-1}$.

In term of binary search tree, this is translated by the fact that the insertion of the $n+1$st key in the tree with n internal nodes is uniform among its $n+1$ leaves.

In this model, the law of the sequence of the underlying (unlabelled) trees is a Markov chain $(\mathcal{T}_n, n \ge 0)$ on **BinTree** defined by $\mathcal{T}_0 = \{✠\}$ and

$$\begin{aligned} \mathcal{T}_{n+1} &= \mathcal{T}_n \cup \{D_{n+1}0, D_{n+1}1\}\,, \\ P(D_{n+1} = u \,|\, \mathcal{T}_n) &= (n+1)^{-1}, \quad u \in \partial\mathcal{T}_n\,, \end{aligned} \tag{5.1}$$

(D_{n+1} is the random node where the $n+1$-st key is inserted). It is a particular case ($\alpha = 1$) of the diffusion-limited aggregation (DLA) on a binary tree, where a constant α is given and the growing of the tree is random with probability of insertion at a leaf u proportional to $\alpha^{-|u|}$ (Aldous-Shields [1], Barlow-Pemantle-Perkins [6]).

To describe the evolution of the BST, two important random variables are the saturation level h_n and the height H_n:

$$h_n = \min\{|u| : u \in \partial\mathcal{T}_n\} \quad , \quad H_n = \max\{|u| : u \in \partial\mathcal{T}_n\} \tag{5.2}$$

which grow logarithmically (see for instance Devroye [14])

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{h_n}{\log n} = c' = 0.3733... \qquad \lim_{n\to\infty} \frac{H_n}{\log n} = c = 4.31107...\,, \tag{5.3}$$

where c' and c are the two solutions of the equation $\eta_2(x) = 1$ where

$$\eta_\lambda(x) := x \log \frac{x}{\lambda} - x + \lambda, \quad x \ge 0\,, \tag{5.4}$$

is the Cramer transform of the Poisson distribution of parameter λ. Function η_2 has its minimum at $x = 2$. It corresponds to the rate of propagation of the insertion depth: $\frac{D_n}{2\log n} \xrightarrow{P} 1$.

A more accurate information on $\mathcal{T}_n$ is provided by the whole profile

$$U_k(n) := \#\{u \in \partial\mathcal{T}_n, |u| = k\} \quad , \quad k \ge 1\,, \tag{5.5}$$

counting the number of leaves of $\mathcal{T}_n$ at each level. Notice that $U_k(n) = 0$ for $k > H_n$ and for $k < h_n$. To get asymptotic results, it is rather natural to code the profile thanks to the so-called polynomial level $\sum_k U_k(n) z^k$, whose degree is H_n.

For $z \notin \frac{1}{2}\mathbb{Z}^- = \{0, -1/2, -1, -3/2, \cdots\}$ let

$$\mathcal{M}_n(z) = \frac{1}{C_n(z)} \sum_{k \geq 0} U_k(n) z^k = \frac{1}{C_n(z)} \sum_{u \in \partial \mathcal{T}_n} z^{|u|} \qquad , \ n \geq 0 \,, \tag{5.6}$$

where

$$C_n(z) = \prod_{k=0}^{n-1} \frac{k+2z}{k+1} = (-1)^n \binom{-2z}{n}, \qquad n \geq 1, \qquad C_0(z) = 1 \,, \tag{5.7}$$

and let $\mathcal{F}_{(n)}$ be the σ-field generated by all the events $\{u \in \mathcal{T}_j\}_{j \leq n, u \in \mathbb{U}}$. Jabbour [13, 21] proved that $(\mathcal{M}_n(z), \mathcal{F}_{(n)})_n$ is a martingale to which, for the sake of simplicity we refer from now as the BST martingale. If $z > 0$, this positive martingale is a.s. convergent; the limit $\mathcal{M}_\infty(z)$ is positive a.s. if $z \in (z_c^-, z_c^+)$, with

$$z_c^- = c'/2 = 0.186..., \quad z_c^+ = c/2 = 2.155... \tag{5.8}$$

and $\mathcal{M}_\infty(z) = 0$ for $z \notin [z_c^-, z_c^+]$ (Jabbour [21]). This martingale is also the main tool to prove that the limit profile has a Gaussian shape (see Theorem 1 in [21]).

5.1.2 Embedding of BST in a continuous time model

The aim of the present paper is to revisit the study of this family of martingales, improving results (in the critical case, on the uniformity of convergence), using either the embedding method or the tilting probability method. It allows to get more complete results on the profile of BSTs.

The idea of embedding discrete models (such as urn models) in continuous time branching processes goes back at least to Athreya-Karlin [4]. It is described for instance in Athreya and Ney ([5], section 9) and it has been recently revisited by Janson [22]. For the BST, various embeddings are mentioned in Devroye [14], in particular those due to Pittel [32], and Biggins [10, 11]. Here, we work with a variant of the Yule process, taking into account the tree (or "genealogical") structure.

Let $(u_t)_{t \geq 0}$ be a Poisson point process taking values in $\mathbb{U}$ with intensity measure $\nu_{\mathbb{U}}$, the counting measure on $\mathbb{U}$. Let $(\mathbb{T}_t)_{t \geq 0}$ be a **BinTree** valued process such that $\mathbb{T}_0 = \{\maltese\}$ and $\mathbb{T}_.$ jumps only when $u_.$ jumps. Let t be a jump time for $u_.$; $\mathbb{T}_t$ is obtained from $\mathbb{T}_{t-}$ in the following way:

if $u_t \notin \partial \mathbb{T}_{t-}$ keep $\mathbb{T}_t = \mathbb{T}_{t-}$ and if $u_t \in \partial \mathbb{T}_{t-}$ take $\mathbb{T}_t = \mathbb{T}_{t-} \cup \{u_t 0, u_t 1\}$.
The counting process $(N_t)_{t \geq 0}$ defined by

$$N_t := \# \partial \mathbb{T}_t \tag{5.9}$$

is the classical Yule (or binary fission) process (Athreya-Ney [5]). In the following, we refer to the continuous-time tree process $(\mathbb{T}_t)_{t \geq 0}$ as the Yule tree process.

We note $0 = \tau_0 < \tau_1 < \tau_2 < ...$ the successive jump times (of $\mathbb{T}_.$),

$$\tau_n = \inf\{t : N_t = n+1\} \,. \tag{5.10}$$

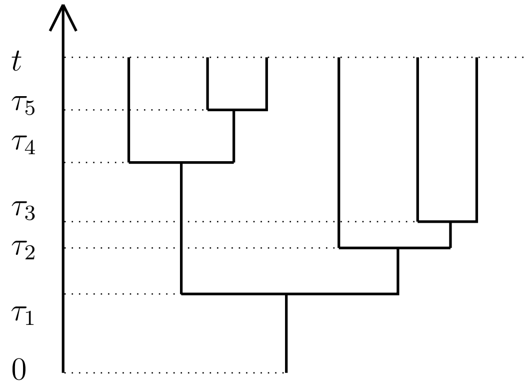

Figure 5.2: Continuous time binary branching process.

5.1.3 Yule process and fragmentation process

This Yule tree process can also be seen as a fragmentation process. We may encode dyadic open subintervals of $[0,1]$ with elements of $\mathbb{U}$. We set $I_{\maltese} = (0,1)$ and for $u = u_1 u_2 ... u_k \in \mathbb{U}$,

$$I_u = \Big(\sum_{j=1}^{k} u_j 2^{-j}, 2^{-k} + \sum_{j=1}^{k} u_j 2^{-j}\Big).$$

With this coding, the evolution corresponding to the previous process is a very simple example of fragmentation process. This idea goes back to Aldous and Shields ([1] Section 7f and 7g). In other words, for $t \geq 0$, $F(t)$ is a finite family of intervals. At time 0, we have $F_0 = (0,1)$. Identically independent exponential $\mathcal{E}(1)$ random variables[2] are associated with each intervals of $F(t)$. Each interval in $F(t)$ splits into two parts (with same size) independently of each other after an exponential time $\mathcal{E}(1)$.

Hence, one has $F(0) = (0,1)$, $F(\tau_1) = ((0,1/2),(1/2,1))$ where $\tau_1 \sim \mathcal{E}(1)$, etc... One can interpret the two fragments I_{u0} and I_{u1} issued from I_u as the two children of I_u, one being the left (resp. right) fragment I_{u0} (resp. I_{u1}), obtaining thus a tree structure. With this interpretation, one observes that when n fragments are present, each of them will split first equally likely. An interval with length 2^{-k} corresponds to a leaf at depth k in the corresponding tree structure.

The following proposition allows to build on the same probability space, the Yule tree process and the BST. This observation was also made in Aldous-Shields [1] section 1, (see also Kingman [23] p.237 and Tavaré [35] p.164 in other contexts).

Lemma 1 *a) The jump time intervals $(\tau_n - \tau_{n-1})_n$ are independent and satisfy:*

$$\tau_n - \tau_{n-1} \sim \mathcal{E}(n) \text{ for any } n \geq 1. \tag{5.11}$$

b) The processes $(\tau_n)_{n\geq 1}$ and $\left(\mathbb{T}_{\tau_n}\right)_{n\geq 1}$ are independent.

[2] $\mathcal{E}(\lambda)$ is the exponential distribution of parameter λ, $\mathcal{U}([0,1])$ is the uniform distribution on $[0,1]$, and $\mathbf{Be}(p)$ is the Bernoulli distribution of parameter p.

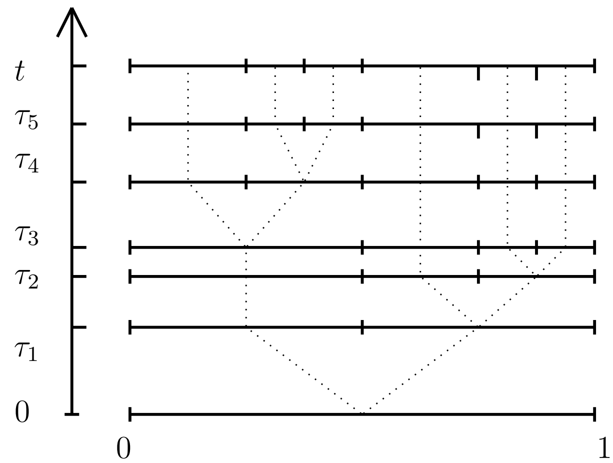

Figure 5.3: Fragmentation and its tree representation.

c) **(embedding)**

$$\left(\mathbb{T}_{\tau_n}\right)_{n\geq 1} \stackrel{\mathcal{D}}{=} \left(\mathcal{T}_n\right)_{n\geq 1} \tag{5.12}$$

where $\stackrel{\mathcal{D}}{=}$ *means equality in distribution.*

Proof: a) and b) are direct consequences of the properties of Poisson processes: a) comes from the definition of the intensity measure, and b) from the independence of jump chain and jump times. c) is clear since the evolution rules of the two Markov chains are the same in both models. ■

A first easy and useful consequence of a) is

$$\mathbb{E}(e^{\tau_n(1-2z)}) = C_n(z)^{-1}\,. \tag{5.13}$$

If we consider only the size of the fragments, the Yule tree process can be seen as a particular case of branching random walk in continuous time: individuals have an $\mathcal{E}(1)$ distributed lifetime, and at their death, they produce children, whose relative positions are distributed according to a point process Z. Individuals do not move during their lives. If we denote the set of individuals alive at time t by $\mathcal{Z}_t$ and for $u \in \mathcal{Z}_t$ the position of individual u by X_u, then the classical family of "additive" martingales, parameterized by θ in $\mathbb{R}$ (sometimes in $\mathbb{C}$) and indexed by $t \geq 0$ is given by

$$m(t,\theta) := \sum_{u\in\mathcal{Z}_t} \exp(\theta X_u - tL(\theta)),$$

where $L(\theta) = E\int e^{\theta x} Z(dx) - 1$ (see [36], [25], and [8] for the fragmentation).

Here, we have $Z = 2\delta_{-\log 2}$, $\mathcal{Z}_t = \partial\mathbb{T}_t$ and $X_u = -|u|\log 2$. For easier use, we set $z = 2^{-\theta}$ and then consider the family of martingales

$$M(t,z) := \sum_{u\in\mathcal{Z}_t} z^{|u|} e^{t(1-2z)}. \tag{5.14}$$

In particular $M(t,1/2)=1$ and $M(t,1)=e^{-t}N_t$. A classical result (see Athreya-Ney [5] or Devroye [14] 5.4) says that

$$\xi := \lim_{t\to\infty} e^{-t}N_t \ \sim \mathcal{E}(1)\,. \tag{5.15}$$

Taking again the (very) particular case $z=1$, we remark that since $\lim_n \tau_n = \infty$ a.s. (see Lemma 1 a)) we get from (5.15)

$$\lim_n ne^{-\tau_n} = \xi \quad \text{a.s.}\,. \tag{5.16}$$

The definitions of the martingales together with the embedding Lemma 1 c) give:

Proposition 1 (martingale connection) *For* $z \notin \frac{1}{2}\mathbb{Z}^-$

$$M(\tau_n, z) = e^{\tau_n(1-2z)}C_n(z)\mathcal{M}_n(z)\,, \tag{5.17}$$

where τ_n *is independent of* $\mathcal{M}_n(z)$.

This connection allows us to transfer known results about the Yule martingales to BST martingales, thus giving a very simple proof of known results (such that in Theorem 1 below) about the BST martingale and also getting much more. In particular, in Theorem 4 2), we give the answer to the question asked in [21], about critical values of z, with a straightforward argument.

5.1.4 Tiltings of the models

We introduce now (and develop in Section 5.3) the tilting or biasing method which allows us to interpret the martingales as Radon-Nikodym derivatives. In order to do that, we need to "enlarge" the probability space [8, 12, 26]. Roughly speaking it consists in marking at random a special "ray" or branch of the tree, called spine, both in the discrete and in the continuous case. It allows to deduce important properties of the population from the behavior of the spine.

For the fragmentation process $(F(t))_{t\geq 0}$ let us denote by $\mathcal{F}_t$ the σ-algebra of the interval fragmentation up to time t and V be a $\mathcal{U}([0,1])$ r.v. independent of the filtration $(\mathcal{F}_t)_{t\geq 0}$. Since $\mathbb{P}(V \in \{k2^{-j}, 0\leq k\leq 2^j, j\in\mathbb{N}, k\in\mathbb{N}\}) = 0$, we may define $\mathbb{P}$-a.s. for every t a unique $S(t)\in\mathbb{U}$ such that $I_{S(t)}$ is an interval of $F(t)$ and $V\in I_{S(t)}$. In other words, $S(t)$ is the element of $\mathbb{U}$ coding the fragment containing V, its length is $2^{-|S(t)|}$ and

$$\mathbb{P}(S(t)=u \mid \mathcal{F}_t) = 2^{-|u|}\ ,\ u\in\partial\mathbb{T}_t \tag{5.18}$$

(we choose a fragment at random with probability equals to its length, it is the classical size-biasing setting). As a consequence of the general theory of homogeneous fragmentations (see Bertoin [7]) or by a direct computation, we see that $(|S(t)|, t\geq 0)$ is an homogeneous Poisson process with parameter 1. In particular, if

$$\mathcal{E}(t,z) := (2z)^{|S(t)|}e^{t(1-2z)} \tag{5.19}$$

then $\mathbb{E}\mathcal{E}(t,z) = 1$. Conditionally on $\widehat{\mathcal{F}}_s = \mathcal{F}_s \vee \sigma(S(r), r \leq s)$, the restriction of the fragmentation $F(.+s)$ to the interval $I_{S(s)}$ is distributed as a rescaling of $F(.)$ by a factor $2^{-|S(s)|}$, which entails that $\big(\mathcal{E}(t,z), \widehat{\mathcal{F}}_t\big)_{t\geq 0}$ is a martingale. By the size biasing scheme (5.18) and the definition (5.14) we get

$$M(t,z) = \mathbb{E}\left[\mathcal{E}(t,z) \mid \mathcal{F}_t\right]. \tag{5.20}$$

Coming back to the discrete time, let $\mathcal{F}_{(n)}$ be the σ-algebra generated by $F(\tau_1), ..., F(\tau_n)$ and let us denote $\mathbf{Spine}_n := S(\tau_n)$ and $s_n := |\mathbf{Spine}_n|$. Applying (5.18) at the $(\mathcal{F}_t, t \geq 0)$ stopping time τ_n, we get for every leaf $u \in \partial\mathcal{T}_n$ (and $k \geq 1$) :

$$\mathbb{P}(\mathbf{Spine}_n = u \mid \mathcal{F}_{(n)}) = 2^{-|u|} \quad , \quad \mathbb{P}(s_n = k \mid \mathcal{F}_{(n)}) = U_k(n)2^{-|k|} \,. \tag{5.21}$$

Let $\widehat{\mathcal{F}}_{(0)}$ be the trivial σ-algebra, and for $n \geq 1$ let $\widehat{\mathcal{F}}_{(n)}$ be the σ-algebra obtained from $\mathcal{F}_{(n)}$ by adjunction of $S(\tau_1), ..., S(\tau_n)$. Let us consider $\mathcal{E}_n(z) := \mathbb{E}\Big[\mathcal{E}(\tau_n, z) \mid \widehat{\mathcal{F}}_{(n)}\Big]$ (with $\mathcal{E}_0(z) := 1$). From (5.13) we have

$$\mathcal{E}_n(z) = (2z)^{s_n}\, C_n(z)^{-1}. \tag{5.22}$$

From the martingale property of $\mathcal{E}(t,z)$ and the definition of $\mathcal{E}_n(z)$ we see that $\big(\mathcal{E}_n(z), \widehat{\mathcal{F}}_{(n)}\big)$ is a martingale. Like in (5.20), we get easily

$$\mathcal{M}_n(z) = \mathbb{E}\left[\mathcal{E}_n(z) \mid \mathcal{F}_{(n)}\right], \tag{5.23}$$

so that the martingales $M(t,z)$ and $\mathcal{M}_n(z)$ are obtained from the "exponential martingales" $\mathcal{E}(z,t)$ and $\mathcal{E}_n(z)$ by projection. All these martingales are precisely the main tool to tilt probabilities. In particular we define $\mathbb{P}^{(2z)}$ on $(\widehat{\mathcal{F}}_t, t \geq 0)$ by

$$\mathbb{P}^{(2z)}_{|\widehat{\mathcal{F}}_t} = \mathcal{E}(t,z)\ \mathbb{P}_{|\widehat{\mathcal{F}}_t}\,, \tag{5.24}$$

which yields by projection on $(\mathcal{F}_t, t \geq 0)$

$$\mathbb{P}^{(2z)}_{|\mathcal{F}_t} = M(t,z)\ \mathbb{P}_{|\mathcal{F}_t}\,. \tag{5.25}$$

If ${}^d\mathbb{P}$ (resp. ${}^d\mathbb{P}^{(2z)}$) is the restriction of $\mathbb{P}$ (resp. $\mathbb{P}^{(2z)}$) to $\vee_n \widehat{\mathcal{F}}_{(n)}$, the discrete versions of the above relations are

$${}^d\mathbb{P}^{(2z)}_{|\widehat{\mathcal{F}}_{(n)}} = \mathcal{E}_n(z)\ {}^d\mathbb{P}_{|\widehat{\mathcal{F}}_{(n)}} \quad , \quad {}^d\mathbb{P}^{(2z)}_{|\mathcal{F}_{(n)}} = \mathcal{M}_n(z)\ {}^d\mathbb{P}_{|\mathcal{F}_{(n)}}\,. \tag{5.26}$$

The probabilities $\mathbb{P}^{(2z)}$ that are given above will have a representation (or an interpretation) further in the paper. In one word, one can say that under $\mathbb{P}^{(2z)}$ the evolution of the fragmentation (or the the one of the Yule tree or BST tree) is biaised. The parameter z serves for a speed-tuner of the spine ($z > 1/2$ corresponds to a speed up and $z < 1/2$ to a slow down).

Let us now explain the content of the paper. In Section 5.2, we explore the direct consequences of embedding. First, we exhibit (on the same space of probability) a family of uniform r.v. attached to the nodes of the tree. These random variables give, for every node u, the limiting proportion of leaves issued from u among those issued from its parent. They are exactly the r.v. called "fictitious" by Devroye in [14] p. 258 in its "backward" construction of a typical realization of $\mathcal{T}_n$ for n fixed. In a second part of Section 5.2, we study the convergence of the BST martingale $\mathcal{M}_n(z)$. For $z > 0$, the embedding method allows to recover very quickly the behavior of the limit $\mathcal{M}_\infty(z)$: positive when $z \in (z_c^-, z_c^+)$, zero when $z \notin [z_c^-, z_c^+]$. In the critical case $z = z_c^\pm$ the behavior was unknown. We prove that $\mathcal{M}_\infty(z_c^\pm) = 0$ a.s. and get the convergence of the derivative. We also give a strong version of the "quicksort" equation.

In Section 5.3, we define the biased models (continuous time and discrete time). After enlarging the space, we prove that the growing tree process can be decomposed into a spine evolution together with the evolution of subtrees issued from nodes of the spine. We follow the way initiated by Lyons, Pemantle, Peres ([26],[27]) and followed by many other authors ([3],[12],[6]). In this study, we use several times the Chinese restaurant model of Dubins and Pitman ([31] p.58). In Section 5.4, we explore the benefits of the tilting method. In a first part, we revisit the behavior of the martingales $\mathcal{M}_n(z)$, giving "conceptual" proofs. In a second part, thanks to this method, we are able to describe the asymptotic behavior of the profile $U_k(n)$ when $k = 2z\log n + o(\sqrt{\log n})$ in the whole range $z \in (z_c^-, z_c^+)$, providing large deviations results around $k = 2\log n$. Previously, the result was known only on a subdomain due to a L^2 method ([13]).

5.2 Some benefits of the embedding method

Let us begin with the study of some meaningful random variables arising as a.s limits and playing an important role in the results of Subsection 5.2.2.

5.2.1 Uniform r.v. in the BST

For every $u \in \mathbb{U}$, let $\tau^{(u)} = \inf\{t : u \in \mathbb{T}_t\}$ the time (a.s. finite) at which u appears in the tree, and for $t > 0$, let

$$\mathbb{T}_t^{(u)} = \{v \in \mathbb{U} \colon uv \in \mathbb{T}_{t+\tau^{(u)}}\}$$

the tree process growing from u. In particular we denote

$$N_t^{(u)} = \#\partial\mathbb{T}_t^{(u)}\,.$$

For $t > \tau^{(u)}$, the number of leaves at time t in the subtree issued from node u is $n_t^{(u)} := N_{t-\tau^{(u)}}^{(u)}$. The branching property and (5.15) give that a.s. for every $u \in \mathbb{U}$

$$\lim_{t\to\infty} e^{-t} N_t^{(u)} = \xi_u \quad , \quad \lim_{t\to\infty} e^{-t} n_t^{(u)} = \xi_u\, e^{-\tau^{(u)}}\,, \tag{5.27}$$

where ξ_u is distributed as ξ i.e. $\mathcal{E}(1)$. Moreover, if u and v are not in the same line of descent, the r.v. ξ_u and ξ_v are independent. Since

$$n_t^{(u)} = n_t^{(u0)} + n_t^{(u1)} \quad \text{and} \quad \tau^{(u0)} = \tau^{(u1)}, \tag{5.28}$$

a small computation yields

$$\frac{n_t^{(u0)}}{n_t^{(u)}} \xrightarrow{a.s.} U^{(u0)} := \frac{\xi_{u0}}{\xi_{u0}+\xi_{u1}}, \qquad \frac{n_t^{(u1)}}{n_t^{(u)}} \xrightarrow{a.s.} U^{(u1)} := 1 - U^{(u0)} = \frac{\xi_{u1}}{\xi_{u0}+\xi_{u1}}, \tag{5.29}$$

which allows to attach a $\mathcal{U}([0,1])$ r.v. to each node of $\mathbb{U}$. In particular we set

$$U := U^{(0)} = \frac{\xi_0}{\xi_0+\xi_1} \tag{5.30}$$

so that

$$\xi := \xi_{\maltese} = e^{-\tau_1}(\xi_0+\xi_1) \quad , \quad \xi_0 = U\xi e^{\tau_1} \quad , \quad \xi_1 = (1-U)\xi e^{\tau_1} \, . \tag{5.31}$$

If $u0$ and $u1$ are brother nodes, we have $U^{(u1)} + U^{(u0)} = 1$. We claim that if the finite set of nodes $v_1, \dots, v_k$ does not contain any pair of brothers, the corresponding r.v. $U^{(v_1)}, \dots, U^{(v_k)}$ are independent. When none of the v_j is an ancestor of another ("stopping line" property) it is a consequence of the branching property. In the general case, it is sufficient to prove that $U^{(u)}$ is independent of $(U^{(v)}, v < u)$. To simplify the reading, let us give the details only for $|u| = 2$, for instance $u = 00$. We have, from (5.28)

$$U^{(00)} = \frac{\xi_{00}}{\xi_{00}+\xi_{01}} \quad , \quad U^{(0)} = \frac{(\xi_{00}+\xi_{01})e^{-\tau^{(00)}+\tau^{(0)}}}{(\xi_{00}+\xi_{01})e^{-\tau^{(00)}+\tau^{(0)}} + (\xi_{10}+\xi_{11})e^{-\tau^{(10)}+\tau^{(1)}}}$$

Actually, from the branching property, ξ_{00} and ξ_{01} are independent of $\xi_{10}, \xi_{11}, \tau^{(00)}, \tau^{(0)}, \tau^{(10)}, \tau^{(1)}$. Moreover since ξ_{00} and ξ_{01} are independent and $\mathcal{E}(1)$ distributed, then $\xi_{00}/(\xi_{00}+\xi_{01})$ and $(\xi_{00}+\xi_{01})$ are independent, which allows to conclude that $U^{(00)}$ and $U^{(0)}$ are independent.

Finally, multiplying along the line of the ancestors of a node u, we get the representation

$$a.s. \lim_{t\to\infty} \frac{n_t^{(u)}}{N_t} = \prod_{v<u} U^{(v)} \, . \tag{5.32}$$

Notice that relation (5.32) gives a strong (which means a.s.) version of the analogy between BST and branching random walks, first given by Devroye [14].

5.2.2 Martingales

For both models a family of martingales plays an essential role: the discrete-time martingale (5.6) in the BST, and the continuous time "additive martingale" (5.14) in the Yule tree. They are closely related by the martingale connection of Proposition 1. Thus, the embedding method is the key tool for proving and enlarging convergence results on the BST martingale (Theorem 4) and its derivative (Theorem 5).

Known results

The following theorem gives a summary of the main properties of the BST martingale, proved in [21] and [13].

Theorem 1 *1) For $z \in (0, \infty)$, the positive martingale $\mathcal{M}_n(z)$ is a.s. convergent when $n \to \infty$ and the limit denoted $\mathcal{M}_\infty(z)$ satisfies*

$$\mathbb{E}\big(e^{-\theta \mathcal{M}_\infty(z)}\big) = \int_0^1 \mathbb{E}\big(e^{-\theta z x^{2z-1} \mathcal{M}_\infty(z)}\big)\mathbb{E}\big(e^{-\theta z(1-x)^{2z-1} \mathcal{M}_\infty(z)}\big)dx\,; \tag{5.33}$$

2) a) for $z \in (z_c^-, z_c^+)$ there exists $p > 1$ such that the L^p convergence holds, and

$$\mathcal{M}_\infty(z) > 0 \quad a.s.\,,$$

b) for $z \notin [z_c^-, z_c^+]$

$$\mathcal{M}_\infty(z) = 0 \ a.s.$$

3) On every compact of $\{z \in \mathbb{C} : |z-1| < \frac{\sqrt{2}}{2}\}$, $\mathcal{M}_n(z)$ and all its z-derivative are a.s. uniformly convergent as $n \to \infty$.

As a consequence of known results for the branching random walks ([8, 9, 36]), we have for the additive martingale:

Theorem 2 *1) For $z \in (z_c^-, z_c^+)$, the positive martingale $M(t, z)$ is a.s. convergent when $t \to \infty$ and the limit denoted by $M(\infty, z)$ satisfies*

$$M(\infty, z) = z e^{(1-2z)\tau_1} \left(M_0(\infty, z) + M_1(\infty, z)\right) \quad a.s. \tag{5.34}$$

where $M_0(\infty, z)$ and $M_1(\infty, z)$ are independent, distributed as $M(\infty, z)$ and independent of τ_1.

2) a) For $z \in (z_c^-, z_c^+)$ there exists $p > 1$ such that the L^p convergence holds, and

$$M(\infty, z) > 0 \quad a.s.\,.$$

b) For $z \in (0, \infty) \setminus (z_c^-, z_c^+)$, then $M(\infty, z) = 0$ a.s..

Notice that the zero limit at the critical points z_c^- and z_c^+ is known in the continuous-time case and not in the discrete-time case.

The derivative

$$M'(t, z) = \frac{d}{dz} M(t, z)$$

is a martingale which is no more positive. It is called the derivative martingale. Its behavior is ruled by the following theorem.

Theorem 3 *1) For $z \in (z_c^-, z_c^+)$, the derivative martingale is convergent a.s. when $n \to \infty$.*

2) a) For $z = z_c^-$, the derivative martingale is convergent a.s. to a finite positive limit $M'(\infty, z_c^-)$ and $\mathbb{E}\big(M'(\infty, z_c^-)\big) = +\infty$.

b) For $z = z_c^+$, the derivative martingale is convergent a.s. to a finite negative limit $M'(\infty, z_c^+)$ and $\mathbb{E}\big(M'(\infty, z_c^+)\big) = -\infty$.

New results

Theorem 4 *1) For $z \in (0,\infty)$ we have a.s.*

a) **(limit martingale connection)**

$$M(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\, \mathcal{M}_\infty(z)\,, \tag{5.35}$$

where $\xi \sim \mathcal{E}(1)$ is defined in (5.15), and independent of $\mathcal{M}_\infty(z)$.

b)

$$\mathcal{M}_\infty(z) = z\left(U^{2z-1}\mathcal{M}_{\infty,(0)}(z) + (1-U)^{2z-1}\mathcal{M}_{\infty,(1)}(z)\right) \tag{5.36}$$

where $U \sim \mathcal{U}([0,1])$ is defined in (5.29), $\mathcal{M}_{\infty,(0)}(z), \mathcal{M}_{\infty,(1)}(z)$ are independent (and independent of U) and distributed as $\mathcal{M}_\infty(z)$.

2) For $z = z_c^\pm$, $\mathcal{M}_\infty(z) = 0$ a.s.

The results on the derivative martingales

$$\mathcal{M}'_n(z) = \frac{d}{dz}\mathcal{M}_n(z)$$

are given in the following theorem, where Ψ the digamma function is defined by

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \lim_n \left(\log n - \sum_{j=0}^{n-1}\frac{1}{x+j}\right). \tag{5.37}$$

Theorem 5 *1) For $z \in [z_c^-, z_c^+]$, $\mathcal{M}'_n(z)$ converges a.s. and its limit $\mathcal{M}'_\infty(z)$ is related to $\mathcal{M}_\infty(z)$ and $M'(\infty,z)$ by*

$$M'(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\left(\mathcal{M}'_\infty(z) + 2\left(\log\xi - \Psi(2z)\right)\mathcal{M}_\infty(z)\right) \qquad a.s. \tag{5.38}$$

where $\xi \sim \mathcal{E}(1)$ is defined in (5.15) and is independent of $\mathcal{M}_\infty(z)$ and $\mathcal{M}'_\infty(z)$. Moreover $\mathcal{M}'_\infty(z)$ satisfies a.s.

$$\begin{aligned} \mathcal{M}'_\infty(z) &= zU^{2z-1}\mathcal{M}'_{\infty,(0)}(z) + z(1-U)^{2z-1}\mathcal{M}'_{\infty,(1)}(z) \\ &+ 2z\left(U^{2z-1}\log U\right)\mathcal{M}_{\infty,(0)}(z) + 2z\left((1-U)^{2z-1}\log(1-U)\right)\mathcal{M}_{\infty,(1)}(z) \\ &+ z^{-1}\mathcal{M}_\infty(z) \end{aligned} \tag{5.39}$$

where $U \sim \mathcal{U}([0,1])$ is defined in (5.29), and the r.v. $\mathcal{M}'_{\infty,(0)}(z)$ and $\mathcal{M}'_{\infty,(1)}(z)$ are independent (and independent of U) and distributed as $\mathcal{M}'_\infty(z)$.

2) a) $\mathcal{M}'_\infty(z_c^-) > 0$ and $\mathcal{M}'_\infty(z_c^+) < 0$ a.s.

b) $\mathbb{E}\left(\mathcal{M}'_\infty(z_c^-)\right) = -\mathbb{E}\left(\mathcal{M}'_\infty(z_c^+)\right) = +\infty$

c) For $z = z_c^{\pm}$, $\mathcal{M}'_\infty(z)$ satisfies the same equation as in Theorem 4 b)

$$\mathcal{M}'_\infty(z) = z\Big(U^{2z-1}\mathcal{M}'_{\infty,(0)}(z) + (1-U)^{2z-1}\mathcal{M}'_{\infty,(1)}(z)\Big) \quad a.s.\,, \tag{5.40}$$

where U, $\mathcal{M}'_{\infty,(0)}(z), \mathcal{M}'_{\infty,(1)}(z)$ are as above. Moreover

$$M'(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\, \mathcal{M}'_\infty(z)\,. \tag{5.41}$$

An easy and remarkable consequence of Theorem 5 1) is obtained in the following corollary, just taking $z = 1$ in (5.38) and (5.39) (remember that $\mathcal{M}_n(1) \equiv 1$). The distribution version (weaker) of (5.43) below is the subject of a broad literature (see for instance Fill, Janson, Devroye, Neininger, Rösler, Rüschendorf [18, 19, 15, 29, 34, 33]) and some properties of the distribution of $\mathcal{M}'_\infty(1)$ remain unknown.

Corollary 2 *We have*

$$M'(\infty, 1) = \xi\left(\mathcal{M}'_\infty(1) + 2\left(\log\xi + \gamma - 1\right)\right) \quad a.s.\,, \tag{5.42}$$

where γ is the Euler constant, and $\mathcal{M}'_\infty(1)$ satisfies the a.s. version of the **quicksort** *equation:*

$$\mathcal{M}'_\infty(1) = U\mathcal{M}'_{\infty,(0)}(1) + (1-U)\mathcal{M}'_{\infty,(1)}(1) + 2U\log U + 2(1-U)\log(1-U) + 1\,, \tag{5.43}$$

where as above, $\mathcal{M}'_{\infty,(0)}(1)$ and $\mathcal{M}'_{\infty,(1)}(1)$ are independent (and independent of U), distributed as $\mathcal{M}'_\infty(1)$ and $U \sim \mathcal{U}([0,1])$.

The following proposition gives an answer to a natural question asked in [13]: what is the optimal domain *in the complex plane* where the BST martingale is L^1-convergent and uniformly convergent? Notice that for $z \in \mathbb{R}$, the notations coincide with those of [21].

Theorem 6 *Let*

$$f(z,q) := 1 + q(2\Re z - 1) - 2|z|^q\ .$$

Let $\mathcal{V}_q = \{z : f(z,q) > 0\}$ and $\mathcal{V} := \cup_{1<q<2}\mathcal{V}_q$.

As $n \to \infty$, $\{\mathcal{M}_n(z)\}$ converges, a.s. and in L^1, uniformly on every compact C of $\mathcal{V}$.

Proofs

In this section we use several times the following lemma.

Lemma 3 *For $z \notin \frac{1}{2}\mathbb{Z}^-$ we have*

$$a)\qquad C_n(z) \sim \frac{n^{2z-1}}{\Gamma(2z)}\,, \tag{5.44}$$

$$b)\qquad a.s.\ \lim_n e^{\tau_n(1-2z)}C_n(z) = \frac{\xi^{2z-1}}{\Gamma(2z)}. \tag{5.45}$$

$$c)\qquad a.s.\ \lim_n \left[\frac{C_n'(z)}{C_n(z)} - 2\tau_n\right] = 2[-\Psi(2z) + \log\xi]\,. \tag{5.46}$$

Proof: a) Use Stirling formula.

b) By (5.44) and (5.16) we get $\lim_n e^{\tau_n(1-2z)} n^{2z-1} = \xi^{2z-1}$, a.s..

c) Use

$$\frac{C_n'(z)}{C_n(z)} = \sum_{j=0}^{n-1} \frac{2}{j+2z},$$

(5.16) and (5.37). ■

Proof of Theorem 4:

1) a) Since $M(t,z)$ converges a.s. when $t \to \infty$, and since $\lim_n \tau_n = \infty$ a.s. we have $\lim_n M(\tau_n, z) = M(\infty, z)$. It remains to apply the martingale connection Proposition 1 and Lemma 3.

b) For $t > \tau_1$ we have the decomposition

$$M(t,z) = ze^{(1-2z)\tau_1}\left[M^{(0)}(t-\tau_1,z) + M^{(1)}(t-\tau_1,z)\right] \tag{5.47}$$

where for $i = 0, 1$

$$M^{(i)}(s,z) = \sum_{u \in \partial \mathbf{T}_s^{(i)}} z^{|u|} e^{s(1-2z)},$$

and $\mathbb{T}^{(i)}$ is defined in Section 5.2.1. Take $t = \tau_n$ in (5.47), condition on the first split time τ_1, apply the branching property, let $n \to \infty$ and apply the limit martingale connection (5.35) to get

$$\frac{\xi^{2z-1}}{\Gamma(2z)}\mathcal{M}_\infty(z) = ze^{(1-2z)\tau_1}\left(\frac{\xi_0^{2z-1}}{\Gamma(2z)}\mathcal{M}_{\infty,(0)}(z) + \frac{\xi_1^{2z-1}}{\Gamma(2z)}\mathcal{M}_{\infty,(1)}(z)\right) \tag{5.48}$$

where ξ_0 and ξ_1 come from section 5.2.1, which yields b) with the help of (5.31).

2) The result for critical points comes directly from the limit martingale connection (5.35) and the analogous known result in continuous time in Theorem 2 2) b). ■

Proof of Theorem 5: 1) Taking derivatives in the martingale connection (Proposition 1) gives

$$M'(\tau_n,z) = \left[\frac{C_n'(z)}{C_n(z)} - 2\tau_n\right] e^{\tau_n(1-2z)} C_n(z)\mathcal{M}_n(z) + e^{\tau_n(1-2z)} C_n(z)\mathcal{M}_n'(z). \tag{5.49}$$

For $z \in [z_c^-, z_c^+]$, let $n \to \infty$ in (5.49). From Lemma 3 and known results in continuous time (Theorem 3) , we get that $\mathcal{M}'(\infty)$ satisfies (5.38).

To prove (5.39), we differentiate (5.47) with respect to z

$$M'(t,z) = (z^{-1} - 2\tau_1)M(t,z) + ze^{(1-2z)\tau_1}\left[M^{(0)'}(t-\tau_1,z) + M^{(1)'}(t-\tau_1,z)\right],$$

and we use the same technique as above: take $t = \tau_n$, let $n \to \infty$, apply (5.38) and its analogs with $(M'^{(i)}, \mathcal{M}^{(i)}, \mathcal{M}'^{(i)}, \xi_i)_{i=0,1}$ instead of $(M', \mathcal{M}, \mathcal{M}', \xi)$, and use (5.31).

For $z = z_c^\pm$, 2) a) and 2) b) are consequences of Theorem 3 2), since $\mathcal{M}_\infty(z_c^\pm) = 0$.

Formula (5.40) of 2) c) is straightforward from (5.39) since $\mathcal{M}_\infty(z_c^\pm) = 0$. Formula (5.41) is (5.38) for $z = z_c^\pm$. ∎

Proof of Theorem 6: Uniform convergence of martingales in the continuous time BRW has been studied by Biggins [9] Theorem 6. See also Bertoin-Rouault [8].

It is possible to give a proof of the uniform a.s. convergence of $\mathcal{M}_n$ directly from these papers. Actually, for the uniform L^1 convergence, we will prove

$$\lim_N \sup_{n\geq N} \mathbb{E} \sup_{z\in C} |\mathcal{M}_n(z) - \mathcal{M}_N(z)| = 0\,. \tag{5.50}$$

Since $(\sup_{z\in C} |\mathcal{M}_n(z) - \mathcal{M}_N(z)|)_{n\geq N}$ is a submartingale, this will imply also the a.s. uniform convergence. From (5.13) and the martingale connection (Proposition 1), we have

$$\mathcal{M}_n(z) - \mathcal{M}_N(z) = \mathbb{E}[M(\tau_n, z) - M(\tau_N, z)|\mathcal{F}_{(n)}]$$

so that taking supremum and expectation we get

$$\mathbb{E} \sup_{z\in C} |\mathcal{M}_n(z) - \mathcal{M}_N(z)| \leq \mathbb{E}\left(\sup_{z\in C} |M(\tau_n, z) - M(\tau_N, z)|\right)\,.$$

Taking again the supremum in n we get

$$\begin{aligned} \sup_{n\geq N} \mathbb{E} \sup_{z\in C} |\mathcal{M}_n(z) - \mathcal{M}_N(z)| &\leq \mathbb{E} \sup_{n\geq N} \left(\sup_{z\in C} |M(\tau_n, z) - M(\tau_N, z)|\right) \\ &\leq \mathbb{E} \sup_{T\geq \tau_N} \left(\sup_{z\in C} |M(T, z) - M(\tau_N, z)|\right)\,. \end{aligned} \tag{5.51}$$

We set

$$\Gamma(t) = \sup_{T\geq t} \left(\sup_{z\in C} |M(T, z) - M(t, z)|\right)\,. \tag{5.52}$$

Since, a.s. the martingale $M(t, z)$ converges uniformly for $z \in C$ we have $\lim_t \Gamma(t) = 0$ a.s.. From equation (11) in the proof of Proposition 3 in [8], $\Gamma(0)$ is in L^1. Since a.s. $\Gamma(t) \leq \Gamma(0)$, using the dominated convergence theorem, we get

$$\lim_t \mathbb{E}\big(\Gamma(t)\big) = 0. \tag{5.53}$$

Now, since a.s. $\lim_n \tau_n = +\infty$, we have $\lim_n \mathbb{E}\big(\Gamma(\tau_n)\big) = 0$ which allows to end the proof. ∎

5.3 Biased models and tilting probability

In this Section, we construct an enlarged probability space and we describe the tools (spine evolution, Chinese restaurant) which will give the key arguments in the proof of Theorems 7 and 9 of Section 5.4.

5.3.1 Construction of biased trees

We call *marked tree* a tree with a distinguished leaf. More precisely let

$$\mathbf{MBinTree} := \{(T, u); T \in \mathbf{BinTree}, u \in \partial T\}.$$

If $\widetilde{T} = (T, u)$ is a marked tree, we say that u is the red leaf of T, that $\{v \in \partial T, v \neq u\}$ is the set of *blue* leaves of T, and that ancestors of u are red nodes, and other internal nodes are blue. We denote by $\partial\widetilde{T}$ the set of leaves of T with their colors.

Let $z > 0$ be a parameter. We define on $\mathbb{U} \times \{red, blue\} \times \{0, 1\}$ a Poisson point process $\tilde{u}_t = (u_t, c_t, \epsilon_t)$ with intensity measure

$$\nu_{\mathbb{U}} \otimes \{2z\delta_{red} + \delta_{blue}\} \otimes \{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\}$$

and we denote by $\mathbb{Q}^{(2z)}$ its law. Let us construct continuous time process $(\widetilde{\mathbb{T}}_t)_{t\geq 0}$ with values in **MBinTree** which starts from

$$\widetilde{\mathbb{T}}_0 = (✠, ✠),$$

such that $\widetilde{\mathbb{T}}_.$ jumps only when $\tilde{u}_.$ jumps. Let t be a jump time for $\tilde{u}_.$; $\widetilde{\mathbb{T}}_t$ is obtained from $\widetilde{\mathbb{T}}_{t-}$ in the following way:

– if $(u_t, c_t) \notin \partial\widetilde{\mathbb{T}}_t$, then $\widetilde{\mathbb{T}}_t = \widetilde{\mathbb{T}}_{t-}$

– if $(u_t, c_t) \in \partial\widetilde{\mathbb{T}}_t$, then the new tree is $\mathbb{T}_t = \mathbb{T}_{t-} \cup \{u_t0, u_t1\}$ and its colors are given by

• if $c_t = blue$, the blue leaf u_t becomes a blue node and two (new) blue leaves u_t0, u_t1 appear.

• if $c_t = red$, the red leaf u_t becomes a red node, two new leaves appear: $u_t\epsilon_t$ which is red and $u_t(1 - \epsilon_t)$ which is blue.

One notes again the successive jump times $(\tilde{\tau}_n)_n$. Once again, $(\tilde{\tau}_n - \tilde{\tau}_{n-1}), n \geq 1$ are independent and exponentially distributed, and

$$\tilde{\tau}_n - \tilde{\tau}_{n-1} \sim \mathcal{E}(n - 1 + 2z).$$

It is clear by construction that the set of red nodes is a branch ; the red branch is called **Spine** and $\mathbf{Spine}_n$ is the red leaf of $\widetilde{\mathbb{T}}_{\tilde{\tau}_n}$. Its length is $s_n = |\mathbf{Spine}_n|$.

In terms of the second construction of the Yule process, we have now two kinds of nodes (blue and red). With each blue node u is associated a clock $\mathcal{E}(1)$, and at its death it gives two blue nodes $u0$ and $u1$. With each red node u is associated a clock $\mathcal{E}(2z)$ and at its death it gives two nodes a red one $u0$ (resp. $u1$) and a blue one $u1$ (resp. u0) with probability 1/2 (resp. 1/2). Ancestor is assumed red.

We can also see this process as a branching process with immigration, as presented in [35] (see also [31] chap. 10 and [17]). The spine is a Poisson process of rate $2z$ and at each jump time of this process begins a new Poisson process of rate 1 independent of the spine process and independent of all other Poisson processes already running.

One can again define a discrete time process

$$\widetilde{\mathcal{T}}_n = \widetilde{\mathbb{T}}_{\tilde{\tau}_n}$$

containing all the tree structure (and the color) of $\widetilde{\mathbb{T}}_{\tau_n}$. The discrete evolution is as follows: $\widetilde{\mathcal{T}}_n$ is a complete binary tree with $2n+1$ nodes, n blue leaves and one red one. To construct $\widetilde{\mathcal{T}}_{n+1}$ we choose the red leaf with probability $2z/(n+2z)$ and each blue one with probability $1/(n+2z)$.

– If the chosen leaf v is blue, $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v1, blue), (v0, blue)\}$.

– If the leaf chosen v is red, we toss a fair coin. We put $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v1, blue), (v0, red)\}$, if the coin is heads and $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v0, blue), (v1, red)\}$ if it is tails.

5.3.2 Tilted probability

We use the change of measure defined in (5.26) that we now recall:

$$ {}^{d}\mathbb{P}^{(2z)}_{|\widehat{\mathcal{F}}_{(n)}} = \frac{(2z)^{s_n}}{C_n(z)}\ {}^{d}\mathbb{P}_{|\widehat{\mathcal{F}}_{(n)}} \tag{5.54} $$

so that, in particular ${}^{d}\mathbb{P}^{(1)} = {}^{d}\mathbb{P}$. We often omit the superscript d for simplicity when no confusion is possible. Proposition 2 below gives an intuitive interpretation of the change of probability done in (5.26).

Proposition 2 *The law of $(\widetilde{\mathcal{T}}_n)_n$ under $\mathbb{Q}^{(2z)}$ is ${}^{d}\mathbb{P}^{(2z)}$. It is called a biased BST and for simplicity we denote by $\widetilde{\mathbb{P}} := {}^{d}\mathbb{P}^{(2z)}$ the biased probability.*

Proof: The dynamics we described above yields the following conditional probabilities. For any blue leaf v, $(v, blue) \in \partial\widetilde{\mathcal{T}}_n$ and

$$ \mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = \mathbf{Spine}_n, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, blue)(v1, blue)|\widetilde{\mathcal{T}}_n) = \frac{1}{n+2z} $$

For the red leaf, $\mathbf{Spine}_n = v$ and

$$ \mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = v0, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, red)(v1, blue)|\widetilde{\mathcal{T}}_n) = \frac{1}{2}\frac{2z}{n+2z} $$

similarly,

$$ \mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = v1, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, blue)(v1, red)|\widetilde{\mathcal{T}}_n) = \frac{1}{2}\frac{2z}{n+2z}\,. $$

Summing up, we have for any colored tree $\widetilde{t}_{n+1}$ with $n+1$ nodes that can be obtained from $\widetilde{\mathcal{T}}_n$ by one insertion

$$ \mathbb{Q}^{(2z)}(\widetilde{\mathcal{T}}_{n+1} = \widetilde{t}_{n+1}|\widetilde{\mathcal{T}}_n) = \frac{z^{s_{n+1}-s_n}}{n+2z} \tag{5.55} $$

and

$$ \mathbb{Q}^{(1)}(\widetilde{\mathcal{T}}_{n+1} = \widetilde{t}_{n+1}|\widetilde{\mathcal{T}}_n) = \frac{(1/2)^{s_{n+1}-s_n}}{n+1}. $$

It is clear that $\mathbb{Q}^{(2z)}/Q(1) = \prod_{j=0}^{n-1} \frac{(2z)^{s_{j+1}-s_j}(j+1)}{j+2z} = (2z)^{S_n}\, C_n(z)^{-1} = \mathcal{E}_n(z)$. Hence, $d\mathbb{Q}^{(2z)}$ is absolutely continuous with respect to $Q^{(1)}$, with the Radon-Nikodym derivative announced in (5.54) (see Lemma 1 and 2 in Biggins [12] for a detailed proof in another context). ■

5.3.3 Spine evolution

Thanks to the previous subsections, it appears that

$$s_n = 1 + \sum_1^{n-1} \epsilon_k \tag{5.56}$$

where $(\epsilon_k)_{k\geq 1}$ are independent and for every $k \geq 1$, $\epsilon_k \sim \mathbf{Be}(\frac{2z}{k+2z})$. In particular,

$$\begin{aligned} \mathbb{E}^{(2z)}(s_n) &= 1 + \sum_1^{n-1} \frac{2z}{k+2z} \\ \mathrm{Var}^{(2z)}\ (s_n) &= \sum_1^{n-1} \frac{2z}{k+2z} - \sum_1^{n-1} \left(\frac{2z}{k+2z}\right)^2 . \end{aligned} \tag{5.57}$$

As $n \to \infty$

$$\begin{aligned} \mathbb{E}^{(2z)}(s_n) &= 2z \log n - 2z\Psi(2z) + o(1) \\ \mathrm{Var}^{(2z)}\ (s_n) &= 2z \log n - 2z\Psi(2z) - 4z^2\Psi'(2z) + o(1) . \end{aligned} \tag{5.58}$$

We can now apply known results on sums of independent r.v. or notice that $s_n - \mathbb{E}^{(2z)}(s_n)$ is a martingale, to get the following asymptotic behavior (see [30]).

Proposition 3 *For any parameter $z > 0$,*

1) *(strong law)*

$$\lim \frac{s_n}{\log n} = 2z\,, \qquad \mathbb{P}^{(2z)} - a.s.. \tag{5.59}$$

2) *(law of the iterated logarithm)* $\mathbb{P}^{(2z)}$*-a.s.*

$$\liminf \frac{s_n - 2z\log n}{2\sqrt{2z \log n \log\log\log n}} = -1\,, \qquad \limsup \frac{s_n - 2z\log n}{2\sqrt{2z \log n \log\log\log n}} = +1. \tag{5.60}$$

3) *(central limit theorem) The distribution of* $\dfrac{s_n - 2z\log n}{\sqrt{2z\log n}}$ *under* $\mathbb{P}^{(2z)}$ *converges to a standard normal distribution* $\mathcal{N}(0,1)$.

4) *(local limit theorem)*

$$\lim_n \sup_k \left| \sqrt{2\pi V_n}\ \mathbb{P}^{(2z)}(s_n = k) - \exp\left(-\frac{(k-\mu_n)^2}{2V_n}\right)\right| = 0 \tag{5.61}$$

where $\mu_n = \mathbb{E}^{(2z)}(s_n)$ *and* $V_n = Var^{(2z)}\ (s_n)$.

5) *(large deviations) The family of distributions of* $(s_n,\ n > 0)$ *under* $\mathbb{P}^{(2z)}$ *satisfies the large deviation principle on* $[0,\infty)$ *with speed* $\log n$ *and rate function* η_{2z} *where the function* η_λ *is defined in (5.4).*

We give more details in Section 4.2.

To study the growing of the biased BST away from the spine, we need to recall the Chinese restaurant model.

5.3.4 Chinese restaurant model (CRM)

Let $\theta > 0$ be a parameter. We recall here the Pitman $(0, \theta)$ Chinese restaurant model (see Pitman [31] p.58). An initially empty restaurant has an unlimited number of tables numbered 1,2,..., each capable of seating an unlimited number of customers. Customers 1,2,... arrive one by one and are seated according to the following:

Person 1 sits at table 1. For $n \geq 1$ assume that n customers have already entered the restaurant, and are seated in some arrangement, with at least one customer at each of the table j, for j from 1 to k, where k is the number of tables occupied by the n first customers to arrive. Let $A_j(n)$ the number of customers on table j at time n. The $n+1$-st customer sits at table j with probability $A_j(n)/(n+\theta)$ for any $j \leq k$. With probability $\theta/(\theta+n)$, the $n+1$-st customer sits on the new table $k+1$.

The sequence

$$A(n) = (A_1(n), A_2(n), \dots)$$

is a Markov chain which describes the evolution of the occupation of the Chinese restaurant, we denote by $\mathbf{CR}^{(\theta)}$ its distribution.

Further we will take $\theta = 2z$. So, for $z > 1/2$ the creation of new tables is encouraged. This has to be compared with the speed-tuner of the spine.

5.3.5 Decomposition of the biased BST along the Spine

For every n, on $\widetilde{\mathcal{T}}_n$ there is a red branch. Each blue node and each blue leaf has some red ancestors. We class the blue leaves of $\widetilde{\mathcal{T}}_n$ according to their highest red ancestors; in other words, let $u_0, u_1, \dots, u_{s_n}$ be the set of red nodes (with $u_0 = \maltese$ and for $i \geq 1$, u_i is the red child of u_{i-1}). We denote by

$$S_i(n) = \{u | u \text{ blue leaf of } \widetilde{\mathcal{T}}_n, u_i \text{ highest red ancestor of } u\}.$$

See Figure 5.4.

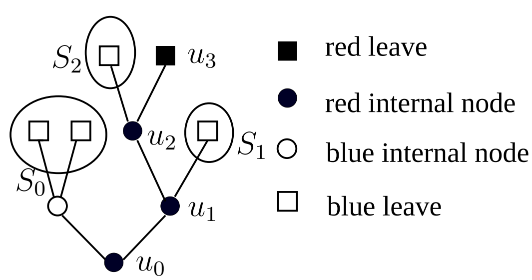

Figure 5.4: A marked tree and the different classes.

We denote by $|S_i(n)| = \operatorname{card} S_i(n)$ and

$$\mathbf{S}(n) = (|S_0(n)|, |S_1(n)|, \dots)$$

the sequence of classes sizes at time n. It satisfies

$$\mathbf{S}(0) = (0, 0, \dots), \quad \mathbf{S}(1) = (1, 0, 0, \dots),$$

and for any n,

$$\sum_{i=0}^{+\infty} |S_i(n)| = n.$$

Recall that s_n is the height of the red leaf. At time n, the class S_{s_n} is the "first empty class". $(\mathbf{S}(k))_{k\geq 0}$ is a Markov chain whose transition at time n can be described as follows:
$a)$ choose class j with probability $|S_j(n)|/(n+2z)$ and set

$$|S_i(n+1)| = |S_i(n)| + \delta_{ij}$$

for $i = 0, 1, ...$ where δ is the Kronecker symbol.
$b)$ choose the red leaf with probability $2z/(n+2z)$ and set

$$\begin{aligned} |S_i(n+1)| &= |S_i(n)| \text{ for } i < s_n\,, \\ |S_{s_n}(n+1)| &= 1\,, \\ s_{n+1} &= s_n + 1\,. \end{aligned} \tag{5.62}$$

Thus

$$A_i(n) \overset{(d)}{=} |S_{i-1}(n)| \text{ for any } i \geq 1.$$

and we may assert

Proposition 4 *Under $\mathbb{P}^{(2z)}$, the Markov chain $\boldsymbol{S}(n), n \geq 0$ is $\mathbf{CR}^{(2z)}$ distributed.*

In [2] p.52, a Chinese restaurant is also considered for the BST, but associated with the insertion node, which does not allow to keep track of the dynamics of the spine.

We now study some conditional evolutions. Let us denote by $\beta_i = \inf\{j \,|\, u_i \in \widetilde{\mathcal{T}}_j\}$ the birth date of node u_i; we have $\beta_0 = 0$, and, for any $l \geq 1$

$$\beta_l = \inf\Big\{k \;\Big|\; s_k = l\Big\}.$$

It is clear that

$$|S_i(n)| > 0 \iff n \geq \beta_{i+1}.$$

Since at time β_{i+1} there are one red leaf, a unique blue leaf on S_i, and $\beta_{i+1} - 1$ blue leaves on others sub-trees, the evolution **conditionally on** β_{i+1} is given by the following relations:

$$\left\{\begin{aligned} |S_i(\beta_{i+1})| &= 1, \\ |S_i(\beta_{i+1}+1)| &= 1 + b_1 \text{ where } b_1 \sim \mathbf{Be}\Big(\frac{1}{\beta_{i+1}+2z}\Big) \\ &\vdots \quad \vdots \\ |S_i(\beta_{i+1}+k)| &= c_{k-1} + b_k \text{ where } b_k \sim \mathbf{Be}\Big(\frac{c_{k-1}}{\beta_{i+1}+k-1+2z}\Big); \end{aligned}\right. \tag{5.63}$$

(conditionally on $|S_i(\beta_{i+1}+j)| = c_j, j = 1, \ldots, k-1$).

In other words, we have the following proposition.

Proposition 5 *Conditionally on β_{i+1}, the distribution of $(|S_i(\beta_{i+1}+k)|, k \geq 0)$ under $\mathbb{P}^{(2z)}$ is the same as $(A_1(k), k \geq 0)$ under $\mathbf{CR}^{(2z+\beta_{i+1}-1)}$.*

Another decomposition will be useful in the rest of the paper. We use the notation $\widetilde{\mathbb{P}}$ for ${}^d\mathbb{P}^{(2z)}$, $\widetilde{\mathbb{E}}$ for the corresponding expectation, with a superscript if we take a conditional one. We denote by $\Xi = \{\beta_i, i \in \mathbb{N}\}$.

Proposition 6 *Under $\widetilde{\mathbb{P}}^{\Xi}$, for i fixed, the process $(S_i(n), n \geq \beta_{i+1})$ has the same distribution as $\mathcal{T}_{W_i(n)}$, that is a (non-biased) BST with $W_i(n)$ leaves where $(W_i(\beta_{i+1}+k), k \geq 0)$ is an inhomogeneous Markov chain on $\{1, 2, \ldots\}$ with initial state $W_i(\beta_{i+1}) = 1$ and the following transition rule:*

- *If $\beta_{i+1}+k \in \Xi$, then $W_i(\beta_{i+1}+k+1) = W_i(\beta_{i+1}+k)$*
- *If $\beta_{i+1}+k \notin \Xi$, then $W_i(\beta_{i+1}+k+1) - W_i(\beta_{i+1}+k) \sim \mathbf{Be}(\frac{W_i(\beta_{i+1}+k)}{\beta_{i+1}+k})$.*

Notice that this evolution of W_i does not give a contradiction with (5.63) since we are conditioning with respect to Ξ which is richer than $\sigma(\beta_{i+1})$.

Proof: The fact that $W_i(n)$ is the distribution size of $S_i(n)$ is a consequence that at times $(\beta_k)_{k>i}$ a new class is created and so no new node arrives on S_i. It remains to show that knowing $W_i(n)$, $S_i(n)$ is distributed as BST with size $W_i(n)$. This comes from the growing rule of the subtree under S_i. Indeed, knowing that "a node arrives" in S_i, this node is inserted uniformly among the leaves already present on S_i, independently from the past. This growing rule is the same as in the classical BST. ■

5.4 Benefits of tilting

We use the method of the tilted probability to revisit the problem of convergence of the family $\mathcal{M}_n(z)$ (subsection 5.4.1) and the convergence of profile (subsection 5.4.2).

The method was initiated by R. Lyons ([26, 27] and developed in several papers involving branching processes or their generalizations ([24, 3, 12, 8]. The main idea consists in changing probability and studying under $\widetilde{\mathbb{P}}$ the spine evolution and the subtrees issued from nodes of the spine. It is a use to call this method "conceptual".

5.4.1 Conceptual proof of convergence of $(\mathcal{M}_n(z), n \geq 1)$

For every $z > 0$, $\mathcal{M}_n(z)$ is a positive $(\mathcal{F}_{(n)}, \mathbb{P})$ martingale and then converges $\mathbb{P}$-a.s. to $\mathcal{M}_\infty(z)$. The L^1 convergence is equivalent to $\int \mathcal{M}_\infty(z) d\mathbb{P} = 1$.

The main argument to decide on this convergence lies on the following lemma. It comes from a classical result of measure theory (the most frequently cited reference is [16] Th. 4.3.3, see also [3, 12]).

Lemma 4 *Fix $z > 0$ and let $\bar{\mathcal{M}}(z) = \limsup_n \mathcal{M}_n(z)$ (notice that $\bar{\mathcal{M}}(z) = \mathcal{M}_\infty(z)$ $\mathbb{P}$-a.s.)*

i) $\mathcal{M}_n(z)^{-1}$ is a $(\mathcal{F}_{(n)}, \widetilde{\mathbb{P}})$ martingale *ii)* $\int \mathcal{M}_\infty(z) d\mathbb{P} = 1$ *if and only if* $\widetilde{\mathbb{P}}(\bar{\mathcal{M}}(z) < \infty) = 1$. *In that case the two measures* $\widetilde{\mathbb{P}}$ *and* $\mathbb{P}$ *are absolutely continuous on* $\mathcal{F}_{(\infty)} = \vee_n \mathcal{F}_{(n)}$ *with density* $\mathcal{M}_\infty(z)$.

iii) $\mathcal{M}_\infty(z) = 0$ $\mathbb{P}$*-a.s. if and only if* $\widetilde{\mathbb{P}}(\bar{\mathcal{M}}(z) = \infty) = 1$. *In that case the two measures* $\widetilde{\mathbb{P}}$ *and* $\mathbb{P}$ *are are mutually singular on* $\mathcal{F}_{(\infty)}$.

Thanks to this dichotomy, we are now able to give alternative proofs of Theorem 1 and Theorem 4 2). For an easier reading we summarize these results in the following theorem.

Theorem 7 *a) If* $z \notin [z_c^-, z_c^+]$ *(subcritical case), then* $\mathbb{P}$*-a.s.* $\lim_n \mathcal{M}_n(z) = 0$.

b) If $z = z_c^\pm$ *(critical case), then* $\mathbb{P}$*-a.s.* $\lim_n \mathcal{M}_n(z) = 0$.

c) If $z \in (z_c^-, z_c^+)$ *(supercritical), then* $\mathbb{P}$*-a.s.* $\lim_n \mathcal{M}_n(z) > 0$ *.*

Subcritical and critical cases (proof of Theorem 7 a)b))

For any $z \geq 0$, we start from

$$\mathcal{M}_n(z) \geq \frac{z^{s_n}}{C_n(z)} \tag{5.64}$$

and we consider the right hand side under $\widetilde{\mathbb{P}}$. From (5.59) and (5.44) we get

$$\lim_n \frac{s_n \log z - \log C_n(z)}{\log n} = \eta_2(2z) - 1\,, \qquad \widetilde{\mathbb{P}} - a.s.$$

In the "subcritical case", i.e. $z \notin [z_c^-, z_c^+]$, we have $\eta_2(2z) > 1$ (see (5.4)) and

$$\lim_n \frac{z^{s_n}}{C_n(z)} = \infty \quad \widetilde{\mathbb{P}} - a.s.. \tag{5.65}$$

In the critical case, $\eta_2(2z) = 1$, we use directly in (5.64) the law of iterated logarithm (5.60) instead of the strong law (5.59). This yields in both cases $\bar{\mathcal{M}}(z) = \limsup_n \mathcal{M}_n(z) = +\infty$, $\widetilde{\mathbb{P}}$- a.s. and by Lemma (4) iii), $\mathcal{M}_\infty(z) = 0$, $\mathbb{P}$- a.s.

Supercritical case (proof of Theorem 7 c) for $z \in (1/2, z_c^+)$)

We will show that for $z \in (1/2, z_c^+)$

$$\liminf_n \mathcal{M}_n(z) < \infty\,, \ \widetilde{\mathbb{P}}\text{- a.s.}\,. \tag{5.66}$$

This is sufficient since by Lemma 4 i) $\mathcal{M}_n(z)^{-1}$ is $\widetilde{\mathbb{P}}$-a.s. convergent. Its limit is nonzero according to (5.66). It will imply that $\mathcal{M}_n(z)$ converges $\widetilde{\mathbb{P}}$- a.s. and allows to conclude with Lemma (4)ii). We stress that for technical reasons, we were able to reach this aim only for $z \in (1/2, z_c^+)$.

Consider $\sigma_n = \sigma(\beta_i \mathbb{1}_{\beta_i \leq n}, i \geq 1)$ the σ-algebra containing the birth date of the red nodes (before time n). By Fatou's lemma, (5.66) is a consequence of the following result.

Proposition 7 *For* $z \in (1/2, z_c^+)$,

$$\limsup_n \widetilde{\mathbb{E}}^{\sigma_n}(\mathcal{M}_n(z)) < +\infty \quad \widetilde{\mathbb{P}}\text{- } a.s..$$

Proof: With the previous decomposition along the spine and by definition of $S_i(n)$, we may write

$$\sum_{u \in \partial \mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} \sum_{u \in S_i(n)} z^{|u|}$$

hence

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in \partial \mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} \widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in S_i(n)} z^{|u|}\,.$$

For every $i \leq s_n - 1$, we have

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in S_i(n)} z^{|u|} = \widetilde{\mathbb{E}}^{\sigma_n} \widetilde{\mathbb{E}}^{\Xi, |S_i(n)|} \sum_{u \in S_i(n)} z^{|u|}\,.$$

From Proposition 6

$$\widetilde{\mathbb{E}}^{\Xi, |S_i(n)|} \sum_{u \in S_i(n)} z^{|u|} = z^i C_{|S_i(n)|}(z)\,,$$

hence

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in \partial \mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} z^i\, \widetilde{\mathbb{E}}^{\sigma_n} C_{|S_i(n)|}(z)\,. \tag{5.67}$$

The main problem is that, knowing σ_n, $|S_i(n)|$ is difficult to handle. Since

$$k \mapsto C_k(z) \text{ is decreasing for } z < 1/2 \text{ and } k \mapsto C_k(z) \text{ is non decreasing for } z \geq 1/2, \tag{5.68}$$

we introduce a new sequence of random variables that will bound $|S_i(n)|$ for the stochastic order.

Recall that if X and Y are two random variables, we say X dominates Y for the stochastic order if for any $x \in \mathbb{R}$, $\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x)$. It implies that for any increasing function g we have $Eg(X) \geq Eg(Y)$.

From Proposition 6, the law of $|S_i(n)|$ conditionally on $\widetilde{\mathbb{P}}^{\Xi}$ is stochastically dominated by the law of $A_1(n - \beta_{i+1})$ under $\mathbf{CR}^{(\beta_{i+1})}$, and a fortiori by the law of $A_1(n)$.

Since $z \geq 1/2$, thanks to (5.68),

$$\widetilde{\mathbb{E}}^{\sigma_n}\left(C_{|S_i(n)|}(z)\right) \leq \mathbf{CR}^{\beta_{i+1}}(C_{A_1(n)}(z))\,. \tag{5.69}$$

Following Barbour & al. [2] page 93 (equations (4.73), (4.74)),

$$\mathbf{CR}^{(\lambda)}(A_1(k) = a) = \frac{\lambda \lambda^{(k-a)}}{\lambda^{(k)}} \frac{(k-1)!}{(k-a)!}$$

where $x^{(n)} = x(x+1)\dots(x+n-1)$ (this is sometimes called the Polya distribution). In the sequel of this proof $2z = \theta$ so that $\theta^{(n)} = C_n(z)n!$. Finally, one obtains (denoting $\beta = \beta_{i+1}$)

$$\begin{aligned} \mathbf{CR}^{(\beta)}(C_{A_1(n)}(z)) &= \frac{\beta(n-1)!}{\beta^{(n)}} \sum_{a=1}^{n} \frac{\theta^{(a)}}{a!} \frac{\beta^{(n-a)}}{(n-a)!} \\ &= \frac{\beta}{n}\left(\frac{(\theta+\beta)^{(n)}}{\beta^{(n)}} - 1 \right), \end{aligned} \tag{5.70}$$

where for the last display we applied Chu-Vandermonde's formula :

$$\sum_{j=0}^{n} \binom{n}{j} x^{(j)} y^{(n-j)} = (x+y)^{(n)} . \tag{5.71}$$

From (5.67), (5.69) and (5.70) we get

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in \partial \mathcal{T}_n} z^{|u|} \leq z^{s_n} + \sum_{i=1}^{s_n - 1} z^i \frac{\beta_{i+1}}{n} \frac{(\theta + \beta_{i+1})^{(n)}}{\beta_{i+1}^{(n)}} . \tag{5.72}$$

Dividing by $C_n(z)$ and setting

$$a_n(\beta, \theta) := \beta \frac{(\theta+\beta)^{(n)}}{\beta^{(n)}} \frac{(n-1)!}{\theta^{(n)}} = \frac{\Gamma(\theta)\Gamma(\theta+\beta+n)\Gamma(\beta+1)\Gamma(n)}{\Gamma(\theta+\beta)\Gamma(\beta+n)\Gamma(\theta+n)},$$

equation (5.72) can be rewritten

$$\widetilde{\mathbb{E}}^{\sigma_n} \mathcal{M}_n(z) \leq \frac{z^{s_n}}{C_n(z)} + \sum_{i=1}^{s_n-1} a_n(\beta_{i+1}, \theta) z^i . \tag{5.73}$$

Since $\lim_{x \to \infty} \frac{x^\theta \Gamma(x)}{\Gamma(x+\theta)} = 1$, one can find $C(\theta) > 0$ such that for every $x \geq 1$

$$\frac{1}{C(\theta) x^\theta} \leq \frac{\Gamma(x)}{\Gamma(x+\theta)} \leq \frac{C(\theta)}{x^\theta} ,$$

which yields

$$a_n(\beta, \theta) \leq C(\theta)^2 \Gamma(\theta) \frac{(\theta+\beta+n)^\theta}{\beta^\theta n^\theta} .$$

For $\beta \leq n$ and $n > \theta$, this gives $a_n(\beta, \theta) \leq C' \beta^{-\theta}$, where C' is a constant depending only on θ. Since $\beta_{i+1} \leq n$ for $i \leq s_n - 1$, this yields, coming back to the notation $\theta = 2z$

$$\widetilde{\mathbb{E}}^{\sigma_n} \mathcal{M}_n(z) \leq \frac{z^{s_n}}{C_n(z)} + C' \sum_{i=1}^{+\infty} (\beta_{i+1})^{1-2z} z^i . \tag{5.74}$$

Since $s_{\beta_{l-1}} \leq l \leq s_{\beta_{l+1}}$ the strong law (5.59) gives

$$\lim_l \frac{\log \beta_l}{l} = \frac{1}{2z} \quad \widetilde{\mathbb{P}}\text{- a.s.}$$

(recall that $\widetilde{\mathbb{P}} = \mathbb{P}^{(2z)}$), hence

$$\lim_n \left((\beta_{i+1})^{1-2z} z^i\right)^{1/i} = e^{(\eta_2(2z)-1)/2z} < 1 \quad \widetilde{\mathbb{P}}\text{- a.s.}$$

(see (5.4)) and the series in the right hand side of (5.74) converges $\widetilde{\mathbb{P}}$-a.s..

For the same reasons, $\lim_n z^{s_n} C_n(z)^{-1} = 0$, $\widetilde{\mathbb{P}}$-a.s. This ends the proof of Proposition 7 and then the proof of Theorem 7. ∎

5.4.2 Convergence of profiles

Random measures and profiles

The profile of the tree $\mathcal{T}_n$ is the sequence

$$\left(U_k(n), k \geq 1\right).$$

Jabbour in [21] introduced the random measure counting the heights of leaves in $\mathcal{T}_n$

$$r_n := \sum_k U_k(n)\delta_k = \sum_{u \in \partial \mathcal{T}_n} \delta_{|u|}.$$

Extreme points of the support of r_n are h_n and H_n. We are interested in the asymptotic behavior of r_n and of its "local" contributions $U_k(n), k \geq 1$. It is related to the behavior of its intensity $\mathbb{E} r_n$ (which is a non-random measure). We may also look at the random measure counting the heights of leaves in the Yule tree:

$$\rho_t = \sum_{u \in \partial \mathbf{T}_t} \delta_{|u|}.$$

As it is clear from (5.3) and as it appears below, the convenient scalings are $(\log n)^{-1}$ for the BST and t^{-1} for the Yule tree process.

Our purpose is, from the one hand to explore some direct links between ρ_t as $t \to \infty$ and r_n as $n \to \infty$, and from the other hand, to give a conceptual proof of the convergence of profiles. A first result concerns the intensity of these measures.

Proposition 8 *a) For $x > 2$*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left(\rho_t(]xt, \infty[)\right) = 1 - \eta_2(x) \tag{5.75}$$

$$\lim_{n\to\infty} \frac{1}{\log n} \log \mathbb{E}\left(r_n(]x \log n, \infty[)\right) = 1 - \eta_2(x). \tag{5.76}$$

b) For $x < 2$, replacing $]xt, \infty[$ (resp. $]x \log n, \infty[$) by $]0, xt[$ (resp. $]0, x \log n[$) the same results hold.

Remark 1
– for $x \in]c', c[$, $\eta_2(x) < 1$, so there are (in mean) about $n^{1-\eta_2(x)}$ leaves of height $\simeq x \log n$. We call this interval $]c', c[$ "supercritical area".
– for $x \in [0, c'[\cup]c, \infty[$, $\eta_2(x) > 1$, so there are (in mean) a very small number of leaves of height $\simeq x \log n$. We call this set "subcritical area".

We call the set $\{c', c\}$ "critical area" .

Proof: Relation (5.75) is easy to obtain, noticing first that by size biasing, for any nonnegative bounded function f,

$$\int f(x)\rho_t(dx) = \mathbb{E}\big[2^{|S(t)|} f(|S(t)|) \mid \mathcal{F}_t\big]$$

so that

$$\mathbb{E} \int f(x)\rho_t(dx) = \mathbb{E}\big[2^{|S(t)|} f(|S(t)|)\big]$$

and then using large deviations for the Poisson process $(|S(t)|, t \geq 0)$.

For the BST, we have similarly

$$\int f(x) r_n(dx) = \mathbb{E}\big[2^{s_n} f(s_n) \mid \mathcal{F}_{(n)}\big]$$

so that

$$\mathbb{E} \int f(x) r_n(dx) = \mathbb{E}\big[2^{s_n} f(s_n)\big] ,$$

and then (5.76) follows using large deviations for (s_n) of Proposition 3.

Notice that the limit in (5.76) is related to (3) of [13]. ■

In the supercritical area we have a.s. convergences:

Theorem 8 *a) For $x \in (2, c)$ a.s.*

$$\lim_{t\to\infty} \frac{1}{t} \log \rho_t(]xt, \infty[) \quad = \quad 1 - \eta_2(x) \tag{5.77}$$

$$\lim_{n\to\infty} \frac{1}{\log n} \log r_n(]x \log n, \infty[) \quad = \quad 1 - \eta_2(x) \tag{5.78}$$

b) For $x \in (c', 2)$ replacing $]xt, \infty[$ (resp. $]x \log n, \infty[$) by $]0, xt[$ (resp. $]0, x \log n[$) the same results hold.

A sharp (non logarithmic) version of relation (5.77) is proved in section 5 of [9] (see also Theorem 5 in [8]). Relation (5.78) is proved by Jabbour [21] in his Theorem 1 (in an equivalent variant for $\nu_n = (n+1)^{-1} r_n$), using Gartner-Ellis theorem. Let us explain shortly how (5.78) can be deduced from (5.77) by embedding.

Taking $t = \tau_n$ in (5.77) we have

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{1}{\tau_n} \log \rho_{\tau_n}(]x\tau_n, \infty[) = 1 - \eta_2(x) \tag{5.79}$$

Now since $\lim ne^{-\tau_n} = \xi$ (cf. (5.16)) we get

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{1}{\log n} \log r_n(]x\tau_n, \infty[) = 1 - \eta_2(x)\,. \tag{5.80}$$

Taking into account that for every $\gamma' > 1 > \gamma$ a.s. there exists n_0 such that for $n > n_0$, $\gamma' x \log n < x\tau_n < \gamma x \log n$, and using the monotonicity of $a \mapsto r_n(]a, \infty[)$ we get the result.

Let us now consider sharp results for the profile. It is well known that

$$\mathbb{E}\big(U_k(n)\big) = \frac{2^k}{n!} S_n^{(k)}$$

where $S_n^{(k)}$ is the Stirling number of the first kind, so from Hwang ([20]), we get, for any $\ell > 0$ as $n \to \infty$ and $k \to \infty$ such that $r = k/\log n \le \ell$

$$\mathbb{E}\, U_k(n) = \frac{(2\log n)^k}{k!\, n\, \Gamma(r)}(1 + o(1)), \tag{5.81}$$

which yields easily

$$\mathbb{E}\, U_k(n) \sim \frac{n^{1-\eta_2(\frac{k}{\log n})}}{\Gamma(\frac{k}{\log n})\sqrt{2\pi k}} \tag{5.82}$$

(see also [13]).

At the level of random variables Jabbour et al. proved in [13] that $\mathbb{P}$ − a.s.

$$\lim_n \frac{U_k(n)}{\mathbb{E} U_k(n)} = \mathcal{M}_\infty(z)\,, \tag{5.83}$$

for $k = 2z\log n + o(\log n)$ and $z \in [0.6,\ 1.4]$. Since their approach laid on L^2 estimations, they guessed that the range $[0.6,\ 1.4]$ may be extended to $I := (1 - 2^{-1/2}, 1 + 2^{-1/2}) = (0.293..., 1.707...)$ which is the maximal interval corresponding to a L^2 convergence of $\mathcal{M}_n(z)$. In the following subsection we extend the validity of the above result to the entire supercritical interval (z_c^-, z_c^+). Our method consists in adapting the proof of the analogous result in the fragmentation model ([8]). Its main interest is that the random limit $\mathcal{M}_\infty(z)$ appears naturally as the usual Radon-Nikodym derivative $d\widetilde{\mathbb{P}}/d\mathbb{P}$, so it is a “conceptual” proof in the Lyons sense ([26]). Although one can obtain a sharper result adapting for instance the Biggins' method ([9]), we prefere to use the conceptual method to illustrate once more its strength.

Main result

Theorem 9 *For $k = 2z\log n + o(\sqrt{\log n})$ and $z \in (z_c^-, z_c^+)$ then*

$$\lim_n \frac{U_k(n)}{\mathbb{E} U_k(n)} = \mathcal{M}_\infty(z)\,,$$

holds in $\mathbb{P}$-probability.

Proof: Let $\widetilde{\mathbb{P}} := \mathbb{P}^{(2z)}$ as defined in Section 5.3 especially by formulas (5.26).

We actually will prove that

$$\lim_n \frac{U_k(n)}{\mathcal{M}_n(z)\mathbb{E}\big(U_k(n)\big)} = 1 \tag{5.84}$$

in $L^2(\widetilde{\mathbb{P}})$, which will entail (5.83). Indeed, the variables $U_k(n)$ and $\mathcal{M}_n(z)$ are $\mathcal{F}_{(\infty)}$-measurable and $\mathcal{M}_n(z)$ converges $\mathbb{P}$-a.s. to $\mathcal{M}_\infty(z)$. Since z is supercritical, the probabilities $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent on $\mathcal{F}_{(\infty)}$, and then $\dfrac{U_k(n)}{\mathbb{E}U_k(n)}$ will converge in $\mathbb{P}$-probability to $\mathcal{M}_\infty(z)$.

To prove (5.84) we first remark that

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)}{\mathcal{M}_n(z)\mathbb{E}(U_k(n))}\right) = \mathbb{E}\left(\frac{U_k(n)}{\mathbb{E}(U_k(n))}\right) = 1 \tag{5.85}$$

so that it is enough to prove that

$$\limsup_n \widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) \le 1\,. \tag{5.86}$$

Using again the change of probability and the size-biasing, especially formula (5.21), we get

$$\begin{aligned}
\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) &= \mathbb{E}\Big(\frac{U_k(n)^2}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\Big)\\
&= \mathbb{E}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{E}\left(\mathbb{1}_{|s_n|=k}2^{|s_n|}|\mathcal{F}_{(n)}\right)\Big]\\
&= \mathbb{E}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{1}_{|s_n|=k}2^{|s_n|}\Big]\\
&= \widetilde{\mathbb{E}}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{1}_{|s_n|=k}C_n(z)z^{-|s_n|}\Big]\,.
\end{aligned} \tag{5.87}$$

Setting

$$B(k,n) = \frac{C_n(z)z^{-k}}{\mathbb{E}(U_k(n))}\,, \quad A(k,n) = \frac{U_k(n)}{\mathbb{E}(U_k(n))}B(k,n)\,, \tag{5.88}$$

the last display of (5.87) becomes

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) = \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{1}_{|s_n|=k}\right)\,. \tag{5.89}$$

The idea is now to replace $A(k,n)$ (resp. $\mathcal{M}_n(z)$) by a similar quantity $\widehat{A}(k,n)$ (resp. $\widehat{\mathcal{M}_n}(z)$), computed with elements above some "dotted line", then apply the Markov property to $\dfrac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{1}_{|s_n|=k}$ and the local central limit theorem to the remaining part of the trajectory of the spine. Consider again a marked BST . Let $(\alpha_n)_{n\in\mathbb{N}}$ any increasing sequence of integers

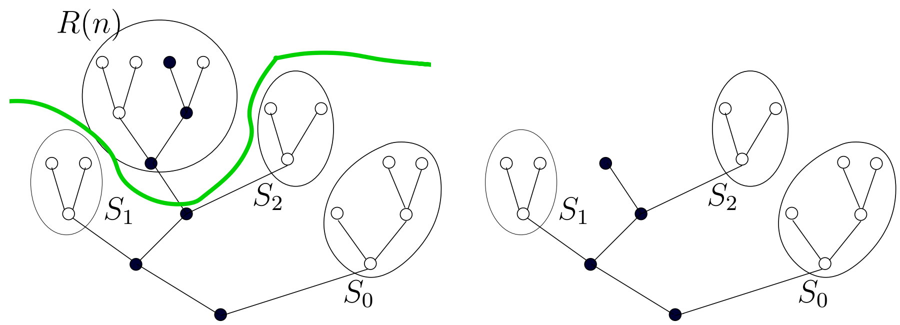

Figure 5.5: A marked tree and the different classes. On the right, what is known in $\mathcal{A}_n$ (and the apparition date of the red points).

and let $\mathcal{A}_n$ be the σ-field

$$\mathcal{A}_n := \sigma\{\beta_0, \beta_1, \dots, \beta_{\alpha_n}, |S_0(n)|, \dots, |S_{\alpha_n}(n)|\};$$

$\mathcal{A}_n$ contains the birth date of the α_n first classes and the number of nodes of each of these classes. Consider $R(n)$, the subtree rooted at u_{α_n} (the red node at level α_n); see Figure 5.5. It contains a red branch and its number of blue leaves is

$$J_n = n - \sum_{i=1}^{\alpha_n} |S_i(n)|.$$

Lemma 5 *a) Let $\theta \in (0, \infty)$. Under $\mathbb{P}^{(\theta)}$ and conditionally on $\mathcal{A}_n$, $R(n)$ is distributed as $\widetilde{\mathcal{T}}_{J_n}$ (under $\mathbb{P}^{(\theta)}$).*

b) Under $\mathbb{P}$ and conditionally on $\{\beta_{\alpha_n} = \beta\}$, the distribution of J_n is that of $A_1(n-\beta)$ under $\mathbf{CR}^{(\beta)}$ for any $\beta \le n-1$.

Proof: a) It is clear that $R(n)$ has J_n nodes. The problem is to show that $R(n)$ has the good "tree structure" distribution. Insertions in the subtree rooted in u_{α_n} occur at times which are not $\mathcal{A}_n$ measurable. But, as a matter of fact, these insertion times are not important. Suppose that at time j an insertion occurs in the subtree rooted in u_{α_n}. At time $j-1$, there were (say) k blue leaves and one red one in this subtree and (say) m blue leaves in the whole tree ($m = j-2$). A simple computation shows that knowing that the insertion occurs in the subtree rooted in u_{α_n}, the insertion occurs on the red node with probability $\theta/(\theta+k)$ and on each blue leave of the subtree with probability $1/(k+\theta)$. These probabilities do not depend on j. The evolution of the tree structure of the subtree is the same as the one of the usual marked tree.

b) It is the result of Proposition 5 with $z = 1/2$. ■

Let us choose $\alpha_n = \lfloor \sqrt{\log n} \rfloor$ and denote by $\widehat{\zeta}_n = \{v \in \partial\mathcal{T}_n : v \in S_i(n) \text{ for some } i \le \alpha_n\}$ the set of leaves below the "dotted line". Let

$$\widehat{\mathcal{M}}_n(z) := \sum_{u \in \widehat{\zeta}_n} \frac{z^{|u|}}{C_n(z)}\ , \quad \widehat{A}(k,n) := \sum_{u \in \widehat{\zeta}_n} \mathbb{I}_{|u|=k} \frac{C_n(z) z^{-k}}{(\mathbb{E}U_k(n))^2}\,.$$

The cost of taking $\widehat{A}$ and $\widehat{M}$ instead of A and M is given by the following lemma.

Lemma 6 *For every $q > 0$,*

$$\mathbb{E}\left(|\widehat{A}(k,n) - A(k,n)|\right) = o((\log n)^{-q})\ ,\ \mathbb{E}\left(|\widehat{\mathcal{M}}_n(z) - \mathcal{M}_n(z)|\right) = o((\log n)^{-q})\,, \tag{5.90}$$

which implies

$$\widetilde{\mathbb{E}}\Big(\Big|\frac{A(k,n)}{\mathcal{M}_n(z)} - \frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\Big|\Big) = o((\log n)^{-q})\,. \tag{5.91}$$

Proof of Lemma 6 We have

$$\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z) = \sum_{u\in\zeta_n - \widehat{\zeta}_n} \frac{z^{|u|}}{C_n(z)}$$

Using Lemma 5 a) with $\theta = 1$ we get

$$\mathbb{E}^{\mathcal{A}_n}\Big(\sum_{u\in\zeta_n-\widehat{\zeta}_n} z^{|u|}\Big) = z^{\alpha_n} C_{J_n}(z)\,.$$

From Lemma 5 b) and formula (5.70) we have for $\beta \le n-1$

$$\mathbb{E}(C_{J_n}(z)|\beta_{\alpha_n} = \beta) = \mathbf{CR}^{(\beta)}(C_{A_1(n-\beta)}) = \frac{\beta(2z+\beta)^{(n-\beta)}}{(n-\beta)\beta^{(n-\beta)}} - \frac{\beta}{(n-\beta)} \tag{5.92}$$

which from (5.70) gives

$$\mathbb{E}\left(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\right) = \frac{z^{\alpha_n}}{C_n(z)}\mathbb{E}\left(\frac{\beta}{(n-\beta)}\frac{(2z+\beta)^{(n-\beta)}}{\beta^{(n-\beta)}} - \frac{\beta}{(n-\beta)}\right)\,, \tag{5.93}$$

and since

$$\frac{\beta(2z+\beta)^{(n-\beta)}}{(n-\beta)\beta^{(n-\beta)}} = \frac{n}{n-\beta}\frac{C_n(z)}{C_\beta(z)}$$

we get

$$\mathbb{E}\left(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\right) = z^{\alpha_n}\mathbb{E}\left(\frac{1}{(n-\beta)}\left(\frac{n}{C_\beta(z)} - \frac{\beta}{C_n(z)}\right)\right)$$

We use here the exponential martingale (introduced in (5.22)) $(\mathcal{E}_k(z), k \ge 1)$ and its stopping time β_{α_n}, so that the above formula becomes

$$\mathbb{E}\Big(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\Big) = 2^{-\alpha_n}\mathbb{E}\left(\mathcal{E}_{\beta_{\alpha_n}}(z)\Pi_n(z)\right) \tag{5.94}$$

where (writing β for β_n)

$$\Pi_n(z) := \frac{1}{n-\beta}\left(n - \beta\frac{C_\beta(z)}{C_n(z)}\right)\,.$$

We write Π_n for $\Pi_n(z)$ and we will prove that Π_n is bounded by a (deterministic) constant which will give

$$\mathbb{E}\Big(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\Big) = \mathcal{O}(2^{-\alpha_n})\,. \tag{5.95}$$

For $2z < 1$ and $\beta < n$ we have $C_\beta(z) > C_n(z)$ (see (5.68) so that $\Pi_n \leq 1$.

When $2z > 1$, we have

$$(n-\beta)\Pi_n = n\Big(1 - \prod_{k=\beta}^{n-1}\Big(1+\frac{2z}{k}\Big)^{-1}\Big)$$

and using the inequality $1 + \frac{2z}{k} \leq \big(1 - \frac{1}{k}\big)^{-2z}$ we get

$$\Pi_n \leq \frac{n}{(n-\beta)(n-1)^{2z}}\left((n-1)^{2z} - (\beta-1)^{2z}\right) \leq \frac{2zn}{n-1}\,,$$

where the last inequality comes from Taylor formula. So, estimation (5.95) holds in any case.

For A we may use the same method. We have, for ever $u \in \mathbb{U}$,

$$\mathbb{I}_{|u|=k} \leq z^{\alpha_n - k} z^{|u|-\alpha_n}\,. \tag{5.96}$$

Adding and taking conditional expectations, we get

$$\mathbb{E}^{\mathcal{A}_n}\left(\sum_{u\in\zeta_n-\widehat{\zeta}_n} \mathbb{I}_{|u|=k}\right) \leq z^{\alpha_n-k} C_{J_n}(z) \tag{5.97}$$

so that

$$\mathbb{E}\left(|\widehat{A}_n(z) - A_n(z)|\right) = \mathcal{O}(2^{-\alpha_n})B(k,n)^2 \tag{5.98}$$

where $B(k,n)$ was defined in (5.88). From (5.81)

$$B(k,n) \sim \sqrt{2\pi k}\; e^{(2z\log n - k)}\left(\frac{2z\log n}{k}\right)^{-k}\,. \tag{5.99}$$

In particular if $k = 2z\log n + o(\sqrt{\log n})$ then

$$B(k,n) \sim \sqrt{2\pi k} \tag{5.100}$$

so that

$$\mathbb{E}\left(|\widehat{A}_n(z) - A_n(z)|\right) = \mathcal{O}(2^{-\alpha_n})k \tag{5.101}$$

which, joined with (5.95) proves the first part of the lemma.

To get (5.91), it is enough to tilt again and use the triangular inequality. ■

Proof of Theorem 9 (end):

From (5.89) and (5.91) we have

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(EU_k(n))^2}\right) = \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{s_n=k}\right) + o((\log n)^{-q}) . \tag{5.102}$$

By conditioning on the dotted line and applying Lemma 5 a), we may replace the above indicator function by $\widetilde{\mathbb{P}}(s_{n'} = k')$ where $n' = n - \beta_{\alpha_n}$ and $k' = k - \alpha_n$. However, to control n' which is random, we first split the main term of (5.102) into

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{s_n=k}\right) \leq \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) + \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right) . \tag{5.103}$$

On the one hand, we will prove further that, for $\gamma_n = \exp(z^{-1}\alpha_n)$,

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) = o(1) . \tag{5.104}$$

On the other hand, the second term of (5.103) becomes

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right) = \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{\beta_{\alpha_n}<\gamma_n}\widetilde{\mathbb{P}}(s_{n'} = k')\right) . \tag{5.105}$$

Taking into account the local central limit theorem (5.61), we may, for any $\varepsilon > 0$, find $r_0 > 0$ such that for $r \geq r_0$

$$\widetilde{\mathbb{P}}(s_r = k') \leq \frac{1+\varepsilon}{\sqrt{2\pi V_r}} \tag{5.106}$$

and since $V_r = 2z\log r + o(\log r)$ we may assume r_0 large enough to ensure $V_r > 2z(1-\varepsilon)\log r$. Choose n_0 such that $n - \gamma_n \geq r_0$ for $n \geq n_0$. It entails

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{\beta_{\alpha_n}<\gamma_n}\widetilde{\mathbb{P}}(s_{n'} = k')\right) \leq \frac{1+\eta}{\sqrt{4\pi z(1-\eta)(\log(n-\gamma_n))}}\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\right) . \tag{5.107}$$

Now, again by (5.90) and (5.88)

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\right) \leq \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\right) + o((\log n)^{-q} = B(k,n) + o((\log n)^{-q} .$$

From (5.100) this gives

$$\limsup_n \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}_n}(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right) \leq \frac{1+\varepsilon}{\sqrt{1-\varepsilon}} ,$$

for any ε. If we admit (5.104) for a while, equations (5.103) and (5.102) lead to (5.86) which ends the proof of the theorem.

It remains to prove (5.104). By (5.91)

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) = \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) + o((\log n)^{-q})\,. \tag{5.108}$$

Now, since $U(k,n) \leq z^{-k}C_n(z)\mathcal{M}_n(z)$ we have $A(k,n) \leq B(k,n)^2\mathcal{M}_n(z)$ and then

$$\widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) \leq B(k,n)^2\ \widetilde{\mathbb{P}}(\beta_{\alpha_n} > \gamma_n) \leq B(k,n)^2\ \widetilde{\mathbb{P}}(s_{\gamma_n} < \alpha_n) \tag{5.109}$$

(for the last inequality see the definition of β).

As said in Proposition 3 5), the family $(s_\ell,\ \ell > 0)$ satisfies under $\widetilde{\mathbb{P}}$ the large deviation principle on $[0,\infty)$ with speed $\log\ell$ and rate function η_{2z} . Therefore, taking $\gamma_n = \exp\frac{\alpha_n}{z}$ we get

$$\limsup \frac{1}{\alpha_n}\log\widetilde{\mathbb{P}}(s_{\gamma_n} < \alpha_n) \leq -\eta_{2z}(z) = -z(1-\log 2) < 0$$

which, joined with (5.100) gives (5.104). This ends the proof of the theorem. ■

Bibliography

[1] D. Aldous and P. Shields. A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields*, 79:509–542, 1988.

[2] R. Arratia, A. Barbour, and S. Tavaré. Logarithmic combinatorial structures: a probabilistic approach. Preprint (book) available at http://www-hto.usc.edu/books/tavare/index.html.

[3] K. Athreya. Change of measures for Markov chains and the L log L theorem for branching processes. *Bernoulli*, 6:323–338, 1999.

[4] K. B. Athreya and S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.*, 39:1801–1817, 1968.

[5] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, New York, 1972.

[6] M.T. Barlow, R. Pemantle, and E.A. Perkins. Diffusion-limited aggregation on a tree. *Probab. Theory Relat. Fields*, 107:1–60, 1997.

[7] J. Bertoin. Homogeneous fragmentation processes. *Probab. Theory Related Fields*, 121(3):301–318, 2001.

[8] J. Bertoin and A. Rouault. Additive martingales and probability tilting for homogeneous fragmentations. Preprint PMA-808 available at http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2003.

[9] J. D. Biggins. Uniform convergence of martingales in the branching random walk. *Ann. Probab.*, 20(1):137–151, 1992.

[10] J. D. Biggins. How fast does a general branching random walk spread? In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 19–39. Springer, New York, 1997.

[11] J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statist. Probab. Lett.*, 32(4):339–342, 1997.

[12] J.D. Biggins and A.E. Kyprianou. Measure change in multitype branching. Preprint available at http://www.shef.ac.uk/ st1jdb/mcimb.html, 2001.

[13] B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *Ann. Appl. Prob.*, 11:1042–1062, 2001.

[14] L. Devroye. Branching processes and their applications in the analysis of tree structures and tree algorithms. In M. Habib et al., editor, *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1998.

[15] L. Devroye, J.A. Fill, and R. Neininger. Perfect simulation from the quicksort limit distribution. *Electronic Communications in Probability*, 5:95–99, 2000.

[16] R Durrett. *Probability: Theory and Exxamples*. Duxbury, Belmont (CA), 1996.

[17] S. Feng and F.M. Hoppe. Large deviation principles for some random combinatorial structures in population genetics and Brownian motion. *The Annals of Prob.*, 8:975–994, 1998.

[18] J.A. Fill and S. Janson. Approximating the limiting quicksort distribution. In *Special Issue of Analysis on Algorithms*, volume 19, pages 376–406, 2001.

[19] J.A. Fill and S. Janson. Quicksort asymptotics. In *Special Issue of Analysis on Algorithms*, volume 44, pages 4–28, 2002.

[20] H.K. Hwang. Asymptotic expansions for the Stirling numbers of the first kind. *J. Combin. Theory Ser. A*, 71(2):343–351, 1995.

[21] J. Jabbour-Hattab. Martingales and large deviations for binary search trees. *Random Structure and Algorithms*, 19:112–127, 2001.

[22] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. Preprint available at http://www.math.uu.se/∼svante/papers/index.html.

[23] J.F.C. Kingman. The coalescent process. *Stochastic Process. Appl.*, 13:235–248, 1982.

[24] T.G. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the kesten-stigum theorem for mutitype branching processes. In P. Jagers K.B. Athreya, editor, *Classical and Modern Branching Processes*, volume 84, pages 181–186. IMA Volumes in Mathematics and its Applications, Springer, 1997.

[25] A. E. Kyprianou. A note on branching Lévy processes. *Stochastic Process. Appl.*, 82(1):1–14, 1999.

[26] R. Lyons. A simple path to Biggins' martingale convergence for the branching random walk. In P. Jagers K.B. Athreya, editor, *Classical and Modern Branching Processes*, volume 84, pages 217–222. IMA Volumes in Mathematics and its Applications, Springer, 1997.

[27] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of L log L criteria for mean behavior of branching processes. *Ann. Probab.*, 23:1125–1138, 1995.

[28] H. Mahmoud. *Evolution of Random Search Trees.* John Wiley, New York, 1992.

[29] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Annals of App. Probab., to appear*, 2003.

[30] V.V. Petrov. *Sums of independent random variables.* Springer Verlag, 1975.

[31] J. Pitman. Cours de Saint-Flour 2002: Combinatorial Stochastic Processes. available at http://stat-www.berkeley.edu/users/pitman/bibliog.html.

[32] B. Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.

[33] U. Rösler. A limit theorem for "quicksort". *RAIRO, Inform. Thor. Appl.*, 25(1):85–100, 1991.

[34] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2):238–261, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

[35] S. Tavaré. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.*, 25(2):161–168, 1987.

[36] K. Uchiyama. Spatial growth of a branching process of particles living in R. *Ann. Probab.*, 10(4):896–918, 1982.