Martingales discrètes et applications à l'analyse d'algorithmes

Brigitte Chauvin

Mars 2002

Université de Versailles-Saint-Quentin
45 Avenue des Etats Unis, F-78035 Versailles Cedex
chauvin@math.uvsq.fr

1 Les martingales discrètes

1.1 Définitions

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Une filtration est une famille croissante de sous-tribus de $\mathcal{A}$.

Définition 1 *Soit $(\mathcal{F}_n)$ une filtration. Une suite de variables aléatoires réelles (X_n) est une $\mathcal{F}_n$-martingale si pour tout n :*

(i) X_n est $\mathcal{F}_n$-mesurable (on dit que la suite (X_n) est adaptée)
(ii) X_n est intégrable : $\mathbb{E} \mid X_n \mid < \infty$
(iii) $\mathbb{E}(X_{n+1}/\mathcal{F}_n) = X_n$, p.s..

Le mot "martingale" vient du cas, au siècle de Pascal, où X_n représente la fortune d'un joueur après la n-ième partie et $\mathcal{F}_n$ représente son information à propos du jeu à ce moment-là. L'égalité (iii) dit que sa fortune espérée après la prochaine partie est la même que sa fortune actuelle. Une martingale est ainsi un jeu équitable.

Définition 2 *Si (iii) est remplacé par*
$\mathbb{E}(X_{n+1}/\mathcal{F}_n) \leq X_n$, p.s. , on obtient une surmartingale, le jeu est défavorable pour le joueur.
Si (iii) est remplacé par
$\mathbb{E}(X_{n+1}/\mathcal{F}_n) \geq X_n$, p.s. , on obtient une sous-martingale.

Remarque : Si l'on n'a pas de filtration sous la main, on peut toujours prendre $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$.

Quelques conséquences :

1) Pour une martingale (X_n), la suite des espérances $(\mathbb{E}X_n)$ est constante, pour une surmartingale, $(\mathbb{E}X_n)$ est décroissante et pour une sous-martingale, $(\mathbb{E}X_n)$ est croissante.

2) Pour tout entier $k \geq 1$, pour une martingale par exemple :
$\mathbb{E}(X_{n+k}/\mathcal{F}_n) = X_n$, p.s..

3) Si (X_n) est une martingale, $\Delta X_n := X_n - X_{n-1}$ s'appelle l'accroissement de X_n. Alors $\mathbb{E}(\Delta X_n/\mathcal{F}_{n-1}) = 0$ et $X_n = X_0 + \sum_{k=1}^n \Delta X_k$. Inversement, on peut se donner des "différences de martingale" ΔX_n, avec la propriété $\mathbb{E}(\Delta X_n/\mathcal{F}_{n-1}) = 0$ et obtenir une martingale en posant $X_n = X_0 + \sum_{k=1}^n \Delta X_k$. Ces deux modes d'exposition sont fréquemment utilisés.

1.2 Comment trouver des martingales ?

1.2.1 Convexité

Si (M_n) est une martingale, et si φ est une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$, alors $\varphi(M_n)$ est une sous-martingale. Par exemple, dès que (M_n) est une martingale, (M_n^2) est une sous-martingale.

1.2.2 Compensateur

Partons d'une suite (X_n), adaptée, d'accroissement $\Delta X_n := X_n - X_{n-1}$, et posons

$$\Delta\widetilde{X_n} := \mathbb{E}(X_n - X_{n-1}/\mathcal{F}_{n-1}) = \mathbb{E}(X_n/\mathcal{F}_{n-1}) - X_{n-1}.$$

A priori, pour une suite quelconque (X_n), cette quantité n'est pas nulle, c'est le "défaut de martingale". Du coup, la suite $(\widetilde{X_n})$ définie par $\widetilde{X_0} = 0$ et

$$\widetilde{X_n} = \Delta\widetilde{X_1} + \cdots + \Delta\widetilde{X_n}$$

est $\mathcal{F}_{n-1}$-mesurable (on dit qu'elle est prévisible), et

$$\begin{aligned}\mathbb{E}(X_n - \widetilde{X_n}/\mathcal{F}_{n-1}) &= \mathbb{E}(X_{n-1} + \Delta\widetilde{X_n} - \widetilde{X_n}/\mathcal{F}_{n-1}) \\ &= X_{n-1} + \mathbb{E}(-\widetilde{X_{n-1}}/\mathcal{F}_{n-1}) \\ &= X_{n-1} - \widetilde{X_{n-1}},\end{aligned}$$

donc la suite $(X_n - \widetilde{X_n})$ est une martingale. On dit que $(\widetilde{X_n})$ est le "compensateur" de (X_n). Le compensateur est l'unique suite prévisible, nulle en 0, telle que $(X_n - \widetilde{X_n})$ soit une martingale (s'il y en avait une autre, la différence serait une martingale prévisible et nulle en 0, c'est-à-dire nulle).

Cas particulier important : la décomposition de Doob :

Si l'on applique la méthode du compensateur ci-dessus à une sous-martingale (X_n), on obtient

$$\Delta\widetilde{X_n} = \mathbb{E}(X_n/\mathcal{F}_{n-1}) - X_{n-1} \geq 0$$

c'est-à-dire que le compensateur est un processus croissant. Autrement dit, toute sous-martingale se décompose de façon unique en la somme d'une martingale M_n et d'un processus croissant noté A_n :

$$X_n = M_n + A_n$$

avec $A_0 = 0$ et
$$A_n = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1}/\mathcal{F}_{k-1}).$$

Dans le cas encore plus particulier où l'on part d'une martingale (M_n), ceci s'applique à la sous-martingale (M_n^2) qui se décompose donc

$$M_n^2 = \text{ martingale} + < M >_n$$

où $< M >_n$ est la notation habituellement utilisée pour le processus croissant de (M_n^2).

1.2.3 Exemple de base : les sommes de variables i.i.d.

Soit (X_n) une suite de variables aléatoires indépendantes et équidistribuées (i.i.d.) de moyenne $m = \mathbb{E}(X)$ et de variance $\sigma^2 = Var(X)$. On s'intéresse au comportement asymptotique de

$$S_n := X_1 + \cdots + X_n.$$

On constate immédiatement que

$$M_n := S_n - n\mathbb{E}(X)$$

est une $\mathcal{F}_n$-martingale. Cherchons son processus croissant :

$$\begin{aligned}
< M >_n &= \sum_{k=1}^{n} \mathbb{E}(M_k^2 - M_{k-1}^2/\mathcal{F}_{k-1}) \\
&= \sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^2/\mathcal{F}_{k-1}) \\
&= \sum_{k=1}^{n} \mathbb{E}((X_k - \mathbb{E}(X)))^2/\mathcal{F}_{k-1}) \\
&= n\sigma^2.
\end{aligned}$$

Finalement, dans ce cas des sommes de variables i.i.d., la décomposition de Doob s'écrit :
$$(S_n - nm)^2 = \text{ martingale } + n\sigma^2.$$

1.2.4 Renormalisation

Supposons que la suite de variables aléatoires (X_n) adaptées vérifie

$$\mathbb{E}(X_n/\mathcal{F}_{n-1}) = A_{n-1}X_{n-1}$$

où A_{n-1} est $\mathcal{F}_{n-1}$-mesurable et différent de 0 presque sûrement. Il suffit alors de renormaliser la suite (X_n) pour obtenir une martingale :

$$Y_n := \frac{X_n}{\prod_{k=0}^{n-1} A_k}$$

est une $\mathcal{F}_n$-martingale.

<u>Exemple : Arbre de Galton-Watson</u>

Prenons le plus simple des processus de branchement : le processus de Galton-Watson, dans lequel un ancêtre à l'instant 0 donne naissance à k individus à l'instant 1 avec probabilité p_k. Puis les individus continuent à se reproduire à des instants discrets, indépendamment les uns des autres et toujours suivant la loi (p_k). On suppose que la moyenne m de cette loi est finie

$$m := \sum_{k\geq 0} k\ p_k\ \ < +\infty.$$

On s'intéresse au comportement asymptotique de

$$Z_n := \sharp \text{ nœuds à la génération } n$$

A tout instant n,

$$Z_n = \sum_{|u|=n-1} N_u$$

où N_u désigne le nombre de descendants de u, de sorte que

$$\mathbb{E}(Z_n/\mathcal{F}_{n-1}) = \sum_{|u|=n-1} \mathbb{E}(N_u/\mathcal{F}_{n-1}) = mZ_{n-1}\ ;$$

en renormalisant, on obtient que

$$M_n := \frac{Z_n}{m^n}$$

est une martingale.

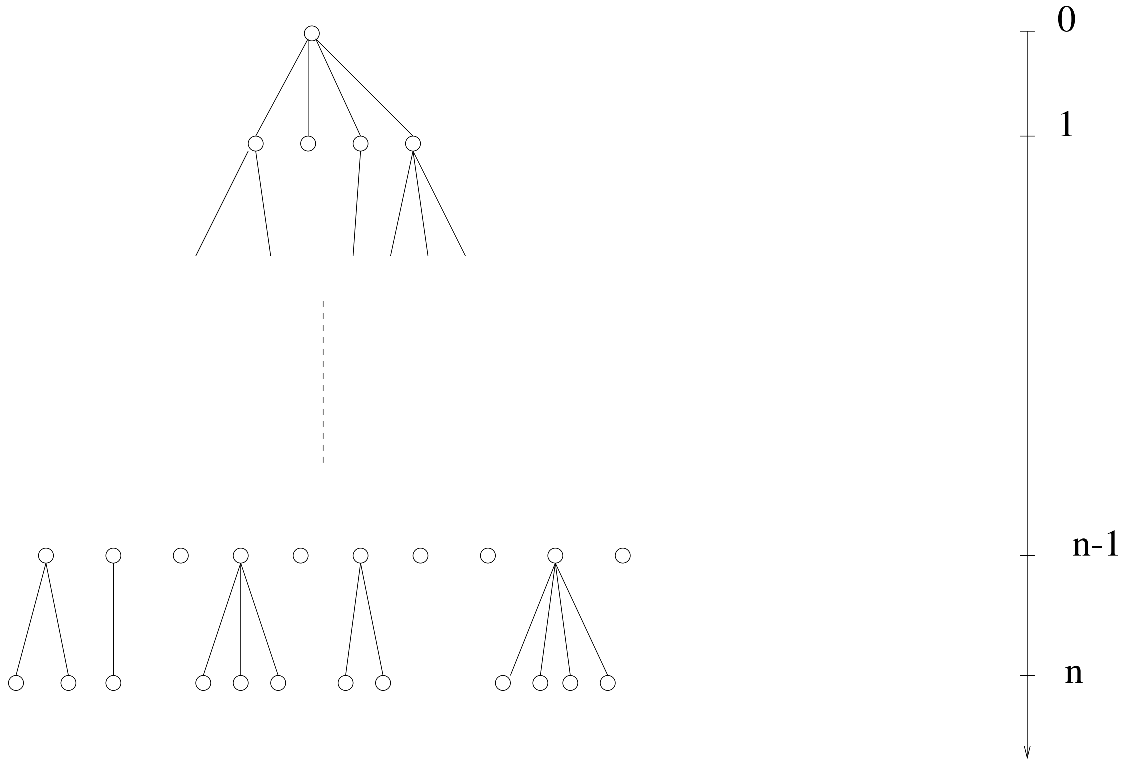


Figure 1 : *arbre de Galton - Watson*

1.2.5 Transformée

Soit (M_n) une martingale et soit (c_n) une suite de variables aléatoires $\mathcal{F}_{n-1}$-mesurables (prévisibles). Les (c_n) peuvent évidemment parfois être réduites à des constantes. Les accroissements de martingale sont $\Delta M_n = M_n - M_{n-1}$ et il est immédiat que

$$Z_n := c_0 M_0 + c_1 \Delta M_1 + \cdots + c_n \Delta M_n$$

est encore une martingale. Z_n s'appelle la "transformée" de M_n.

1.2.6 Chaines de Markov

Partons d'une chaine de Markov (X_n) à temps discret, à valeurs dans un espace d'états S dénombrable et de matrice de transition $P = (p_{i,j})$:

$$\mathbb{P}(X_{n+1} = j / X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j / X_n = i) = p_{i,j}$$

Soit $\psi : S \longrightarrow S$ une fonction harmonique au sens où

$$\forall i \in S, \qquad \sum_{j \in S} p_{i,j} \psi(j) = \psi(i).$$

Alors $(\psi(X_n))_{n \in \mathbb{N}}$ est une martingale pour la filtration associée à (X_n), car

$$\begin{aligned}
\mathbb{E}(\psi(X_{n+1})/X_1, X_2, \ldots, X_n) &= \mathbb{E}(\psi(X_{n+1})/X_n) \\
&= \sum_{j \in S} \mathbb{E}(\psi(j) \mathbb{1}_{\{X_{n+1} = j\}} / X_n) \\
&= \sum_{j \in S} \psi(j) \mathbb{P}(X_{n+1} = j / X_n) \\
&= \sum_{j \in S} p_{X_n, j} \psi(j) \\
&= \psi(X_n).
\end{aligned}$$

Dans le cas un peu plus général où ψ est un vecteur propre de la matrice de transition P, pour la valeur propre λ, c'est-à-dire

$$\forall i \in S, \qquad \sum_{j \in S} p_{i,j} \psi(j) = \lambda \psi(i),$$

alors

$$\mathbb{E}(\psi(X_{n+1})/X_1, X_2, \ldots, X_n) = \lambda\psi(X_n),$$

et en renormalisant, on obtient que

$$M_n := \frac{\psi(X_n)}{\lambda^n}$$

est une martingale.

1.3 Règles d'arrêt

Définition 3 *Un temps d'arrêt T pour la filtration $(\mathcal{F}_n)$ est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$ telle que pour tout entier n, l'événement $\{T \leq n\}$ est $\mathcal{F}_n$-mesurable (c'est la même chose que de demander que pour tout n, $\{T = n\}$ soit $\mathcal{F}_n$-mesurable).*

Si T est un temps d'arrêt, la tribu "arrêtée" à T est

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty, \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

Exemple 1 : $T =$ constante

Exemple 2 : les temps d'atteinte ou temps de premier passage d'une suite adaptée (X_n) par un borélien : soit B un borélien, posons

$$T := \inf\{n \in \mathbb{N}, X_n \in B\};$$

alors

$$\{T = n\} = \{X_1 \notin B, \ldots, X_{n-1} \notin B, X_n \in B\} \in \mathcal{F}_n.$$

On se demande dans la suite si la propriété de martingale peut passer à des instants aléatoires ayant la propriété de temps d'arrêt ?

Théorème 1 *Premier théorème d'arrêt*

(i) Si (X_n) est une martingale et si T est un temps d'arrêt, alors

$$Y_n := X_{T \wedge n}$$

est une martingale, qui s'appelle la martingale arrêtée au temps T.

(ii) Si (X_n) est une martingale et si S et T sont deux temps d'arrêt bornés tels que $S < T$, alors

$$\mathbb{E}(X_T/\mathcal{F}_S) = X_S$$

c'est-à-dire que l'on a la propriété de martingale à des instants aléatoires.

L'hypothèse "temps d'arrêt bornés" est très forte, nous allons chercher à la relaxer, après avoir vu comment convergent les martingales.

Ce théorème vaut également pour les sous-martingales et les surmartingales.

1.4 Inégalités

Les inégalités suivantes concernent le maximum d'une suite de variables aléatoires, et sont appelées inégalités maximales de Doob.

Proposition 1 *Soit (X_n) une sous-martingale positive telle que $\mathbb{E}(X_0) < \infty$, alors*

$$\mathbb{P}(\max_{k \leq n} X_k \geq \lambda) \leq \frac{\mathbb{E}(X_n)}{\lambda}.$$

Quelques remarques

1) La démonstration consiste à introduire T, le temps d'atteinte de λ et à appliquer l'inégalité de Markov à $X_{T \wedge n}$.

2) Si l'on dispose seulement d'une sous-martingale de signe quelconque, il est possible d'appliquer la proposition à $X_n^+ = sup(0, X_n)$ qui est encore une sous-martingale.

3) Si (X_n) est une martingale de signe quelconque, la proposition s'applique à $\mid X_n \mid$ qui est une sous-martingale. Le même argument appliqué au carré d'une martingale conduit au corollaire suivant :

Corollaire 1 *Soit (X_n) une martingale de carré intégrable. Alors*

$$\mathbb{P}(\max_{k \leq n} X_k \geq \lambda) \leq \frac{\mathbb{E}(X_n^2)}{\lambda^2}.$$

1.5 Convergence

Les définitions suivantes ne sont pas universelles.

Définition 4 *On dit qu'une suite de variables aléatoires (X_n) est intégrable (respectivement de carré intégrable) si et seulement si*

$$\mathbb{E}(\mid X_n \mid) < \infty \quad (\textit{ respectivement, } \mathbb{E}(X_n^2) < \infty).$$

On dit qu'une suite de variables aléatoires (X_n) est <u>*bornée dans L^p*</u> *$(p > 0)$, si et seulement si*

$$\sup_n \mathbb{E}(\mid X_n \mid^p) < \infty.$$

On dit qu'une suite de variables aléatoires (X_n) est <u>*équi-intégrable*</u> *ou* <u>*uniformément intégrable*</u> *si et seulement si*

$$\sup_n \mathbb{E}(\mid X_n \mid \mathbb{1}_{\{|X_n|>a\}}) \longrightarrow 0$$

quand $a \to +\infty$.

1.5.1 Convergence dans L^2

Théorème 2 *Convergence L^2 des martingales*
Toute martingale (X_n) bornée dans L^2 converge dans L^2.
Toute sous-martingale positive (X_n) bornée dans L^2 converge dans L^2.

La démonstration de ce théorème repose sur la décomposition de Doob.

Ce théorème est simple à utiliser et il a de nombreuses applications : par exemple, la martingale de l'arbre de Galton-Watson surcritique converge dans L^2, dès que la loi de reproduction a un second moment fini.

1.5.2 Convergence presque sûre

Théorème 3 *Convergence p.s. des martingales. Théorème de Doob.*
Toute sous-martingale (X_n) vérifiant $\sup_n \mathbb{E}(X_n^+) < \infty$ converge p.s. vers une variable aléatoire X_∞ et $X_\infty \in L^1$.

Ce théorème admet de nombreux sous-produits, comme

- Toute sous-martingale (X_n), bornée dans L^1, converge p.s. vers une variable aléatoire X_∞ et $X_\infty \in L^1$.
- Toute martingale (X_n), bornée dans L^1, converge p.s. vers une variable aléatoire X_∞ et $X_\infty \in L^1$.

et l'un de ces corollaires est particulièrement simple et efficace :

Corollaire 2 *Toute surmartingale positive (X_n) converge p.s. vers une variable aléatoire X_∞, $X_\infty \in L^1$ et $\mathbb{E}(X_\infty) \leq \liminf \mathbb{E}(X_n)$ (par Fatou).*

<u>Preuve</u> : regarder $-X_n$.

1.5.3 Convergence dans L^1

Attention : les hypothèses "(X_n) bornée dans L^1" et "(X_n) converge p.s. vers une limite $X_\infty \in L^1$" ne suffisent pas à assurer que X_n converge dans L^1. Il y a de nombreux contre-exemples, les martingales exponentielles de la section 2.1 en sont un.

Théorème 4 *Convergence L^1 des martingales*
Soit (X_n) une martingale. Les trois assertions suivantes sont équivalentes :
(i) (X_n) converge dans L^1 (vers une variable aléatoire $X_\infty \in L^1$)
(ii) (X_n) est bornée dans L^1 et $\mathbb{E}(X_\infty/\mathcal{F}_n) = X_n$.
(iii) (X_n) est uniformément intégrable.

Une martingale vérifiant l'une de ces propriétés est dire régulière. On aura en particulier $\mathbb{E}(X_\infty) = \mathbb{E}(X_0)$.

Il arrive que l'on n'ait pas convergence dans L^2 et que la convergence L^1 soit suspectée mais difficile à obtenir via le théorème précédent (c'est le cas pour des processus de branchement et pour les arbres binaires de recherche par exemple). Une convergence L^p pour $p \in]1, 2[$ pourra alors être utile (outre son intérêt propre).

Corollaire 3 *Convergence $L^p, p > 1$ des martingales*
Toute martingale (X_n) bornée dans L^p pour $p > 1$ converge dans L^p (et aussi en probabilité et p.s. par le théorème de Doob).

Preuve : on montre que $| X_n |^p$ est uniformément intégrable.

1.5.4 Temps d'arrêt et martingales régulières

Définition 5 *Temps d'arrêt régulier.*
On dit qu'un temps d'arrêt T est régulier pour la martingale M_n quand la martingale arrêtée $M_{T\wedge n}$ est régulière.

Cette nouvelle notion permet de raffiner le premier théorème d'arrêt, en afflaiblissant les conditions sur le temps d'arrêt :

Théorème 5 *Second théorème d'arrêt*
Soit (M_n) une martingale dans L^1 (mais pas nécessairement régulière), soient T_1 et T_2 deux temps d'arrêt avec T_2 régulier. Alors

$$\mathbb{E}(M_{T_2}/\mathcal{F}_{T_1}) = M_{T_1} \quad sur\ \{T_1 \leq T_2\}$$

Son corollaire est évidemment utile en pratique :

Corollaire 4 *Soit (M_n) une martingale et soit T un temps d'arrêt régulier. Alors (identité de Wald)*

$$\mathbb{E}(M_n) = \mathbb{E}(M_T)$$

1.6 Théorème central limite

Dans l'exemple de base où la martingale considérée est la somme de v.a. (X_i) i.i.d.,

$$M_n = S_n - n\mathbb{E}(X),$$

le théorème central limite classique s'applique. Dans une variante de l'exemple de base, la martingale considérée est la somme de v.a. (X_i) indépendantes mais pas nécessairement de même loi. Dans ce cas (voir par exemple [14]), le théorème central limite est valide sous une condition de type "condition de Lindeberg". Il est aussi valide sous une condition un peu plus forte, de moment d'ordre $2+\delta$, dite "condition de Lyapunov".

Le cas des martingales est complètement analogue au cas des sommes de v.a. indépendantes et c'est le processus croissant de la martingale qui joue le rôle de la variance (cf [6, 8]).

Théorème 6 *Théorème central limite pour les martingales, version Lindeberg*

Soit (M_n) une martingale centrée et de carré intégrable. Sous les deux conditions

$$\textit{(i)}\quad \frac{<M>_n}{\mathbb{E}(M_n^2)} \longrightarrow \Gamma \textit{ en probabilité}$$

où Γ est une v.a. positive, finie p.s. et

(ii) (condition dite de Lindeberg)

$$\forall \varepsilon > 0,\ \frac{1}{\mathbb{E}(M_n^2)} \sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^2 \mathbb{1}_{\{|M_k - M_{k-1}| > \varepsilon\sqrt{\mathbb{E}(M_n^2)}\}} / \mathcal{F}_{k-1}) \longrightarrow 0 \textit{ en probabilité}$$

Alors

$$\frac{M_n}{\sqrt{\mathbb{E}(M_n^2)}} \longrightarrow \mathcal{N}(0, \Gamma).$$

On note $\mathcal{N}(0,\Gamma)$ pour un mélange de lois normales, au sens où on dira qu'une v.a. Z a pour loi $\mathcal{N}(0,\Gamma)$ si et seulement si sa fonction caractéristique est donnée par

$$\mathbb{E}(e^{-itZ}) = \mathbb{E}(\exp -\frac{1}{2}\Gamma t^2).$$

La version plus simple du théorème central limite pour les martingales, sous condition de moment $2+\delta$ est :

Théorème 7 *Théorème central limite pour les martingales, version Lyapunov*

Soit (M_n) une martingale centrée et de carré intégrable. Sous les deux conditions

$$\textit{(i)}\quad \frac{< M >_n}{\mathbb{E}(M_n^2)} \longrightarrow \Gamma \textit{ en probabilité}$$

où Γ est une v.a. positive, finie p.s. et

(ii) (condition dite de Lyapunov)

$$\frac{1}{(\mathbb{E}(M_n^2))^{1+\delta/2}} \sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^{2+\delta}) \longrightarrow 0 \textit{ en probabilité}$$

Alors

$$\frac{M_n}{\sqrt{\mathbb{E}(M_n^2)}} \longrightarrow \mathcal{N}(0,\Gamma).$$

Dans les cas particuliers suivants, les accroissements de martingale vérifient à chaque fois une condition plus forte que la condition de Lindeberg et le théorème central limite s'appliquera :

(ii) remplacé par $\max_k \frac{\mid M_k - M_{k-1} \mid}{\sqrt{\mathbb{E}(M_n^2)}} \longrightarrow 0$ en probabilité.

(ii) remplacé par $\frac{\mathbb{E}(\max_k (M_k - M_{k-1})^2)}{\mathbb{E}(M_n^2)}$ uniformément bornée.

(ii) remplacé par $\mid M_k - M_{k-1} \mid$ uniformément bornés (très fort).

Enfin, le théorème suivant est une version affaiblie, mais simple à utiliser

Théorème 8 *Soit (M_n) une martingale centrée et de carré intégrable. Supposons qu'il existe une constante K telle que pour tout n,*

$$\mathbb{E}(\mid M_{n+1} - M_n \mid^{2+\delta} / \mathcal{F}_n) \leq K \ .$$

Si la suite $(< M >_n /n)$ *converge en probabilité vers une constante* σ^2*, alors* $\frac{M_n}{\sqrt{n}}$ *converge en loi vers une variable gaussienne* $\mathcal{N}(0, \sigma^2)$.

2 Application aux martingales exponentielles

2.1 Une famille de martingales non régulières

Soit (X_i) une suite de variables aléatoires indépendantes et équidistribuées (i.i.d.) à valeurs réelles, de loi μ qui ne soit pas une mesure de Dirac. On suppose qu'il existe au moins sur un voisinage ouvert de 0, la transformée de Laplace de μ et on appelle l son logarithme :

$$e^{l(u)} := \int_{\mathbb{R}} e^{ux}\ d\mu(x)\ .$$

On s'intéresse au comportement asymptotique de

$$S_n := X_1 + \cdots + X_n.$$

dont on va voir qu'il est très lié aux martingales exponentielles

$$M_n(u) := \exp\left[uS_n - nl(u)\right]\ .$$

On peut comprendre heuristiquement le comportement de $M_n(u)$: quand n est grand, par la loi des grands nombres, $\frac{S_n}{n} \sim \mathbb{E}X$, et comme les moments de X se lisent avec la transformée de Laplace : $l'(0) = \mathbb{E}X$, donc $S_n \sim nl'(0)$, ce qui, en remplaçant dans $M_n(u)$ donne

$$M_n(u) \sim \exp\left[nu(l'(0) - \frac{l(u)}{u})\right]\ .$$

Comme $l(u)$ est convexe, la quantité $l'(0) - \frac{l(u)}{u}$ est négative, ce qui indique que $M_n(u)$ converge exponentiellement vite vers 0. La démonstration rigoureuse de la proposition suivante repose effectivement sur la convexité de $l(u)$.

Proposition 2 *Pour tout réel u appartenant à l'ouvert de définition de $l(u)$,*

$$\boxed{M_n(u) := \exp\left[uS_n - nl(u)\right]}$$

est une martingale positive, d'espérance 1, qui converge p.s. vers 0. Ce n'est pas une martingale régulière.

Démonstration : exercice

Exercice

1) La fonction $u \mapsto f(u) := \exp\left[ux - nl(u)\right]$, pour x et n fixés, est analytique sur V, voisinage de 0 où $l(u)$ est analytique. Son développement est

$$\exp\left[ux - nl(u)\right] = \sum_{k=0}^{\infty} \frac{u^k}{k!} f_k(n, x) \ .$$

Montrer que pour tout entier $k \geq 0$, la suite $(f_k(n, S_n))_n$ est une $\mathcal{F}_n$-martingale intégrable.

2) Regarder ce que l'on obtient pour $k = 1$ et $k = 2$.

3) Montrer que

$$Y_n := \int_{\mathbb{R}} \exp\left[uS_n - nl(u)\right] \, du$$

est encore une martingale positive donc p.s. convergente vers une v.a. finie p.s. Qu'obtient-on dans le cas particulier où les X_i suivent des lois normales centrées réduites ?

2.2 Branching random walks

On trouve aussi des martingales exponentielles dans les marches aléatoires branchantes (branching random walks), puisque des v.a. i.i.d s'ajoutent le long des branches de l'arbre. Rappelons que dans ce processus, un ancêtre (noté $\emptyset$) se trouve en 0 (ou ailleurs) à l'instant 0. Ses enfants forment la première génération et leurs positions (ainsi que leur nombre N) sont données par un processus ponctuel Z sur $\mathbb{R}$. Ensuite, chaque particule u donne naissance, indépendamment des autres particules et du passé, à des enfants selon un processus ponctuel copie de Z.

L'espace de probabilité considéré est celui des arbres marqués par les déplacements γ_u des particules. On définit la position X_u de la particule u par

$$X_u = X_\emptyset + \gamma_{i_1} + \gamma_{i_1 i_2} + \cdots + \gamma_{i_1 \ldots i_n}$$

pour $u = i_1 i_2 \ldots i_n, \qquad i_j \in \mathbb{N}^*$.

Si l'on appelle maintenant μ la mesure d'intensité du processus ponctuel Z, c'est-à-dire que pour toute fonction f mesurable positive

$$\mathbb{E}(Z(f)) = \mathbb{E}(\sum_{j=1}^{N} f(X_j)) = \int f(x)\mu(dx) \ ,$$

supposons que sa transformée de Laplace soit définie sur un voisinage de 0 et notons-la $m(\theta)$:

$$m(\theta) = \mathbb{E}(\sum_{j=1}^{N} e^{\theta X_j}) = \int e^{\theta x}\mu(dx) = \mathbb{E}(Z(e^{\theta\cdot})),$$

son logarithme est toujours noté $l(\theta)$.

$$l(\theta) = \log m(\theta)$$

Théorème 9 *(Kingman, 1975)*
Pour tout θ,

$$W_n(\theta) := \sum_{|u|=n} \exp\left[\theta X_u - nl(\theta)\right]$$

est une $\mathcal{F}_n$-martingale positive, d'espérance 1. *Elle converge presque sûrement vers $W(\theta)$ avec a priori $\mathbb{E}(W(\theta)) \leq 1$ (par le corollaire 2).*

Remarquons qu'en faisant $\theta = 0$, on retrouve le processus de Galton-Watson sous-jacent.

C'est l'étude plus fine de cette martingale (cf [2, 4]), qui permettra d'obtenir des résultats sur la particule la plus à gauche dans ce processus. La connection de Devroye entre un arbre binaire de recherche et une marche aléatoire branchante fournira la convergence presque sûre de la hauteur d'un arbre binaire de recherche.

2.3 Utilisation de martingales exponentielles pour l'étude des queues de distribution de S_n. Identité de Wald.

Pour un réel a donné, l'objectif est d'étudier $\mathbb{P}(S_n \geq a)$. L'idée conductrice est d'appliquer le second théorème d'arrêt (théorème 5) à la martingale exponentielle $M_n(u)$ pour le temps d'arrêt ν_a qui est le temps d'atteinte de a

$$\nu_a := \inf\{n \in \mathbb{N}, S_n \geq a\} \ .$$

Pour cela, il faut voir si ce temps d'arrêt est régulier.

Lemme 1 *Pour tout u tel que $l'(u) > 0$, le temps d'arrêt ν_a est régulier pour la martingale $M_n(u)$.*

Démonstration

La démonstration repose sur un changement de probabilité classique en théorie des grandes déviations.

On doit montrer que $M_{n\wedge\nu_a}(u)$ (dont on sait déjà que c'est une martingale par le premier théorème d'arrêt) est une martingale régulière, c'est-à-dire qu'elle converge dans L^1. Ecrivons

$$M_{n\wedge\nu_a}(u) = M_n(u)1\!\!1_{\{\nu_a > n\}} + M_{\nu_a}(u)1\!\!1_{\{\nu_a \le n\}}$$

et montrons que chacun des deux termes converge dans L^1.

Pour le second terme, il suffit de montrer par convergence dominée que

$$\mathbb{E}(M_{\nu_a}(u)1\!\!1_{\{\nu_a < \infty\}}) < \infty\ .$$

Mais sur l'événement $\{\nu_a \le n\}$, par le premier théorème d'arrêt, la propriété de martingale de $M_{n\wedge\nu_a}(u)$ donne

$$\text{sur } \{\nu_a \le n\},\ M_{\nu_a}(u) = M_{n\wedge\nu_a}(u) = \mathbb{E}(M_n/\mathcal{F}_{\nu_a})$$

et donc

$$\mathbb{E}(M_{\nu_a}(u)1\!\!1_{\{\nu_a \le n\}}) = \mathbb{E}(1\!\!1_{\{\nu_a \le n\}}\mathbb{E}(M_n/\mathcal{F}_{\nu_a})) \le \mathbb{E}(M_n) = \text{Constante.}$$

Pour le premier terme, on va montrer qu'il converge vers 0 dans L^1. Comme la martingale est positive, on va montrer que

$$\mathbb{E}(M_n(u)1\!\!1_{\{\nu_a > n\}}) \longrightarrow 0\ .$$

Ecrivons

$$\begin{aligned}
\mathbb{E}(M_n(u)1\!\!1_{\{\nu_a > n\}}) &= \mathbb{E}(e^{u(X_1+\cdots+X_n)-nl(u)}1\!\!1_{\{\forall m<n, S_m<a\}}) \\
&= \int_{\mathbb{R}^n} e^{ux_1-l(u)}d\mu(x_1)\ldots e^{ux_n-l(u)}d\mu(x_n)1\!\!1_{\{\max\limits_{m<n} S_m\ <\ a\}}
\end{aligned}$$

d'où l'idée de changer de probabilité en posant pour tout u dans le bon voisinage de 0

$$d\mu_u(x) = e^{ux-l(u)}d\mu(x)$$

de sorte que

$$\begin{aligned}
\mathbb{E}(M_n(u)1\!\!1_{\{\nu_a>n\}}) &= \int_{\mathbb{R}^n} d\mu_u(x_1)\dots d\mu_u(x_n)1\!\!1_{\{\max_{m<n} S_m<a\}} \\
&= \mathbb{P}_u(\max_{m<n} S_m < a)
\end{aligned}$$

où $S_m = X_1 + \cdots + X_m$ pour des v.a. (X_i) i.i.d. de loi μ_u. Par la loi des grands nombres, quand $n \to \infty$,

$$\frac{S_n}{n} \longrightarrow \mathbb{E}_u(X) = l'(u) \ p.s.$$

donc si $l'(u) > 0$, S_n tend vers $+\infty$, $\mathbb{P}_u$-p.s. et la probabilité ci-dessus tend vers 0. □

Puisque d'après ce lemme, le temps d'arrêt ν_a est régulier, la martingale est d'espérance constante égale à 1, y compris au temps ν_a. On a ainsi démontré

Proposition 3 *Pour tout u tel que $l'(u) > 0$, le temps d'arrêt ν_a est régulier pour la martingale $M_n(u)$ et on a l'identité de Wald*

$$\mathbb{E}M_{\nu_a}(u) = 1$$

et comme cette martingale converge presque sûrement vers 0, l'identité de Wald s'écrit aussi

$$\mathbb{E}[M_{\nu_a}(u)1\!\!1_{\nu_a<\infty}] = 1$$

Commentaire

Dans les cas où l'on a $S_{\nu_a} = a$, l'identité de Wald s'écrit

$$\mathbb{E}\exp\left[uS_{\nu_a} - \nu_a l(u)\right] = \mathbb{E}\exp\left[ua - \nu_a l(u)\right] = 1 ,$$

ce qui fournit une formule exacte pour $\mathbb{E}(e^{-\nu_a l(u)})$, c'est-à-dire pour la transformée de Laplace du temps d'arrêt ν_a.

3 Martingales à horizon fini et inégalité d'Azuma

L'objectif est de montrer des inégalités de concentration de variables aléatoires X_n autour de leur moyenne $\mathbb{E}X_n$. Pour cela on va construire une martingale artificiellement, seulement pour les instants $0, 1, \ldots, n$ (on parle de martingale à horizon fini), de sorte que X_n soit la martingale à l'instant n et $\mathbb{E}X_n$ soit la martingale à l'instant 0. Puis on utilisera le théorème de concentration suivant pour les martingales ([12] est une référence agréable sur ce sujet).

Théorème 10 *Inégalité d'Hoeffding-Azuma*

Soit M_n une martingale, donnée par les accroissements de martingale ΔM_n ;

$$M_n = M_0 + \Delta M_1 + \cdots + \Delta M_n \ .$$

Supposons qu'il existe une suite de constantes $(c_n)_{n\in\mathbb{N}}$ telle que

$$\forall k \in \mathbb{N}, \qquad | \Delta M_k | \leq c_k \qquad \textit{presque sûrement} \tag{1}$$

Alors, pour tout réel t,

$$\mathbb{P}(| M_n - M_0 | \geq t) \leq 2 \ \exp\big[- \frac{t^2}{2\sum_{k=1}^{n} c_k^2} \big]$$

Exercice

Appliquer le théorème précédent pour montrer que pour une suite de variables aléatoires i.i.d. de loi uniforme sur $[0,1]$, on a l'inégalité de concentration :

$$\mathbb{P}(| S_n - \frac{n}{2} | > t) \leq 2e^{-\frac{2t^2}{n}}$$

Voyons sur un exemple d'application comment la méthode annoncée plus haut peut fonctionner.

Une application au bin-packing

Appelons B_n le nombre de boîtes de taille 1 nécessaires pour ranger n objets de taille $X_1, X_2, \ldots, X_n$ i.i.d. de loi uniforme sur $[0,1]$. On souhaite étudier la variable aléatoire B_n. On sait par exemple que $\frac{B_n}{n}$ converge presque sûrement vers une constante. On cherche une inégalité de concentration de B_n autour de sa moyenne. Pour cela, on construit

$$\forall i = 0, 1, \ldots, n \qquad Y_i^{(n)} := \mathbb{E}(B_n / \mathcal{F}_i) \ .$$

Alors $(Y_i^{(n)})_{i=0,1,\dots,n}$ est une $\mathcal{F}_i$-martingale et

$$Y_n^{(n)} = B_n$$

$$Y_0^{(n)} = \mathbb{E}(B_n) \ .$$

Il suffit donc de montrer que les accroissements de cette martingale (Y_i) vérifient la condition (1) pour en déduire grâce au théorème d'Azuma une majoration de

$$\mathbb{P}(\mid Y_n^{(n)} - Y_0^{(n)} \mid > t) = \mathbb{P}(\mid B_n - \mathbb{E}(B_n) \mid > t) \ .$$

Pour cela, remarquons que si l'on note $B_n(i)$ le nombre de boîtes nécessaires pour ranger les objets sauf X_i, on a toujours

$$B_n(i) \leq B_n \leq B_n(i) + 1$$

ce qui, en passant aux espérances conditionnelles par rapport à $\mathcal{F}_{i-1}$ et $\mathcal{F}_i$ donne

$$\mathbb{E}(B_n(i)/\mathcal{F}_{i-1}) \leq Y_{i-1} \leq \mathbb{E}(B_n(i)/\mathcal{F}_{i-1}) + 1$$
$$\mathbb{E}(B_n(i)/\mathcal{F}_i) \leq Y_i \leq \mathbb{E}(B_n(i)/\mathcal{F}_i) + 1 \ .$$

Mais par définition de $B_n(i)$, les membres de gauche sont égaux, les membres de droite aussi. Donc

$$\mid Y_i - Y_{i-1} \mid \leq 1$$

ce qui est une condition de type (1) avec les constantes toutes égales à 1. L'inégalité d'Azuma fournit ainsi

$$\mathbb{P}(\mid Y_n^{(n)} - Y_0^{(n)} \mid > t) = \mathbb{P}(\mid B_n - \mathbb{E}(B_n) \mid > t) \leq 2e^{-\frac{t^2}{2n}} \ .$$

4 Application aux arbres binaires de recherche

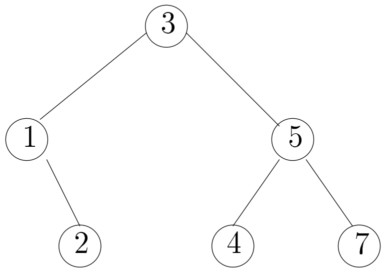
Une référence sur le sujet est le livre de Mahmoud [10]. Rappelons d'abord quelques généralités sur Quicksort et les arbres binaires de recherche.

4.1 Généralités

Un **arbre binaire de recherche** est un arbre binaire dont chaque nœud interne est muni d'une "clé" (d'une étiquette, d'une marque, qui est dans un premier temps un entier) de telle sorte qu'à chaque nœud de l'arbre, toutes les clés du sous-arbre droit sont plus grandes que toutes les clés du sous-arbre gauche.

La définition (récursive) d'un **arbre binaire de recherche** est aussi la suivante : donnons-nous $x_1, x_2, \dots x_n$ réels distincts. L'arbre binaire de recherche (abrégé abr) est l'arbre binaire dans lequel
- x_1 est à la racine
- le sous-arbre gauche est l'abr associé à $\{x_2, \dots x_n\} \cap]-\infty, x_1[$
- le sous-arbre droit est l'abr associé à $\{x_2, \dots x_n\} \cap]x_1, +\infty[$

Exemple : trier 3, 1, 5, 4, 7, 2



Remarque : la lecture de l'arbre en ordre symétrique donne la liste triée.

Insertion : si l'on veut insérer une (n+1)ème donnée x_{n+1}, on la compare à la racine, puis à la clé racine du sous-arbre gauche (ou droit), etc... on ne la compare pas à toutes les clés mais seulement le long d'une branche, jusqu'à l'insérer sur un noeud externe. Appelons D_{n+1} (qui est donc un coût) le niveau d'insertion de la (n+1)ème donnée.

Modèle probabiliste : à tout ensemble de n données distinctes $x_1, x_2, \dots x_n$, on associe un arbre binaire de recherche T_n, à n noeuds internes et n+1 noeuds externes. Le modèle que nous étudions dans la suite, c'est-à-dire la loi de probabilité sur les arbres est l'image (par cette association) de la **loi uniforme sur les permutations** de n objets. Dans ce modèle, les arbres (T_n) ont même loi que ceux construits en choississant $x_1, x_2, \ldots x_n$ selon une densité uniforme sur l'intervalle $[0,1]$. Dans ce modèle, l'insertion d'une (n+1)-ième clé a la même probabilité d'être effectuée à chacun des $n+1$ nœuds externes de T_n.

4.2 Répartition des nœuds externes, largeur d'un abr.

4.2.1 Mesure empirique. Polynôme de niveau

Outre la hauteur et le niveau de saturation, si l'on veut connaître plus finement la répartition des noeuds par niveau dans l'arbre binaire de recherche, on introduit naturellement

$$U_k(n) := \text{le nombre de noeuds externes au niveau } k \text{ dans l'arbre } T_n.$$

On pourrait travailler de façon analogue avec les noeuds internes et introduire

$$V_k(n) := \text{le nombre de noeuds internes au niveau } k \text{ dans l'arbre } T_n.$$

ainsi que

$$\begin{aligned} Z_k(n) &:= \text{le nombre total de noeuds au niveau } k \text{ dans l'arbre } T_n \\ Z_k(n) &= U_k(n) + V_k(n). \end{aligned}$$

Comme il y a $n+1$ noeuds externes dans l'arbre T_n de taille n, la mesure empirique de répartition des noeuds externes est

$$\boxed{\nu_n := \sum_{k=1}^{+\infty} \frac{U_k(n)}{n+1} \, \delta_{\{k\}}.}$$

Comme la $(n+1)$-ième insertion se fait uniformément sur les $n+1$ noeuds externes de l'arbre T_n, la loi du niveau d'insertion, ou profondeur D_n est donnée par

$$\mathbb{P}(D_{n+1} = k \mid T_n) = \frac{U_k(n)}{n+1} = \nu_n(k), \tag{2}$$

et en prenant l'espérance

$$\mathbb{P}(D_{n+1} = k) = \mathbb{E}(\frac{U_k(n)}{n+1}) = \mathbb{E}(\nu_n(k))$$

autrement dit la loi de D_{n+1} est ce qu'on appelle la mesure d'intensité (déterministe) du processus ponctuel ν_n, par conséquent **les résultats en moyenne sur la mesure ν_n donneront des résultats en loi sur D_n.**

Le comportement en moyenne des $U_k(n)$ est connu :

Théorème 11 *(Lynch, 1965)*

$$\mathbb{E}U_k(n) = \frac{2^k}{n!}\begin{bmatrix} n \\ k \end{bmatrix}$$

où $\begin{bmatrix} n \\ k \end{bmatrix}$ sont les nombres de Stirling de première espèce, c'est-à-dire que $\begin{bmatrix} n \\ k \end{bmatrix}$ est le coefficient de x^k dans le développement de $x(x+1)\ldots(x+n-1)$. Ces nombres vérifient la relation de récurrence

$$\begin{bmatrix} n+1 \\ k \end{bmatrix} = \begin{bmatrix} n \\ k-1 \end{bmatrix} + n\begin{bmatrix} n \\ k \end{bmatrix} .$$

Corollaire 5 *La loi de D_n est donnée par*

$$\mathbb{P}(D_{n+1} = k) = \frac{2^k}{(n+1)!}\begin{bmatrix} n \\ k \end{bmatrix} .$$

Le comportement p.s. des $U_k(n)$ peut, lui, se comprendre synthétiquement avec le "polynôme de niveau" qui est défini maintenant.

Définition 6 *On appelle* **polynôme de niveau** *d'un arbre binaire de recherche T_n, le polynôme défini pour tout paramètre $z \in \mathbb{C}$ par*

$$\boxed{W_n(z) := \sum_{k=0}^{+\infty} U_k(n)\ z^k}$$

Il s'agit bien d'un polynôme, car pour $k > 1 + h_n, U_k(n) = 0$. Bien entendu, c'est une variable aléatoire, puisque les $U_k(n)$ sont aléatoires.

Les résultats en moyenne sur le polynôme de niveau se déduisent facilement des résultats en moyenne sur les $U_k(n)$:

$$\mathbb{E}W_n(z) = \sum_{k=0}^{+\infty} \mathbb{E}U_k(n)\ z^k$$

et d'après le théorème 11,

$$\begin{aligned}
\mathbb{E}W_n(z) &= \frac{1}{n!}\sum_{k=0}^{+\infty} 2^k z^k \begin{bmatrix} n \\ k \end{bmatrix} \\
&= \frac{1}{n!} 2z(2z+1)\dots(2z+n-1) \\
\mathbb{E}W_n(z) &= \prod_{j=0}^{n-1} \frac{j+2z}{j+1} \qquad (3)
\end{aligned}$$

La loi de la profondeur est alors immédiate, car si $D_n(z)$ est sa série génératrice

$$D_n(z) = \sum_{k\geq 0} \mathbb{P}(D_n = k) z^k = \mathbb{E}(z^{D_n}) \ ,$$

comme $\mathbb{P}(D_{n+1} = k) = \frac{1}{n+1}\mathbb{E}U_k(n)$ (cf. equation (2)), on obtient

$$D_{n+1}(z) = \frac{1}{n+1}\mathbb{E}W_n(z) = \frac{1}{(n+1)!}\prod_{j=0}^{n-1}(j+2z)$$

qui est une expression assez explicite de la série génératrice de la profondeur d'insertion.

Pour obtenir davantage, c'est-à-dire des résultats p.s. sur le polynôme de niveau, nous allons décrire son évolution dans le temps, et c'est là qu'une martingale va apparaître.

4.3 Résultats p.s. sur le polynôme de niveau : une martingale

La relation évidente

$$U_k(n+1) - U_k(n) = -\mathbb{1}_{\{D_{n+1}=k\}} + 2\ \mathbb{1}_{\{D_{n+1}=k-1\}}$$

permet d'avoir une relation de récurrence sur le "polynôme de niveau" :

$$\begin{aligned}\mathbb{E}(W_{n+1}(z) \mid T_n) &= \sum_{k\geq 0} z^k \Big[U_k(n) + 2\mathbb{P}(D_{n+1} = k-1 \mid T_n) - \mathbb{P}(D_{n+1} = k \mid T_n)\Big] \\ &= \frac{n+2z}{n+1}\, W_n(z)\ ,\end{aligned}$$

ce qui signifie que le polynôme de niveau convenablement renormalisé est une martingale.

Théorème 12 *pour tout nombre complexe $z \in \mathbb{C}$,*

$$M_n(z) := \frac{W_n(z)}{\mathbb{E}(W_n(z))} = \frac{W_n(z)}{\prod_{j=0}^{n-1} \frac{j+2z}{j+1}}$$

est une $\mathcal{F}_n$-martingale.

(i) Elle converge p.s. pour tout z réel positif.

(ii) Elle converge dans L^1 sur l'intervalle réel $]c'/2; c/2[$ (les constantes c et c' étant celles du théorème de Devroye sur la hauteur des arbres binaires de recherche). Elle converge vers 0 en dehors de l'intervalle $[c'/2; c/2]$.

(iii) Elle converge dans L^2 sur la boule $B(1, \frac{1}{\sqrt{2}})$ de $\mathbb{C}$.

Idées de démonstration

La propriété de martingale vient du calcul ci-dessus, par renormalisation.

(i) La convergence presque sûre dans $\mathbb{R}_+$ est celle de toute martingale positive.

(ii) La convergence dans L^1 est plus difficile à voir. Elle est obtenue en bornant $M_n(z)$ dans L^p pour $p > 1$, de façon assez analogue à la méthode utilisée pour un processus de branchement spatial. Les détails sont dans [9].

(iii) La convergence L^2 pour z dans $\mathbb{C}$ s'obtient en calculant la covariance de la martingale et avec un peu d'analyse complexe. La convergence L^1 (et L^p) est un problème ouvert dans $\mathbb{C}$. Les détails se trouvent dans [3]. □

Commentaire : Globalement, cette étude tire parti de l'égalité

$$\boxed{W_n(z) = \mathbb{E}(W_n(z)) M_n(z)}$$

où $M_n(z)$ est une martingale. L'égalité ci-dessus est très parlante : en effet, elle permet de séparer l'étude du pôlynome de niveau en deux parties beaucoup plus simples ; une partie déterministe ($\mathbb{E}(W_n(z))$) donnée exactement par (3) et dont on connaît l'asymptotique ; et une partie aléatoire qui a la propriété de martingale. Tout l'alea est concentré dans cette partie martingale et l'étude de la convergence est facilitée.

Par la suite, des résultats de type théorème central limite et grandes déviations sur la mesure ν_n peuvent être obtenus, ainsi que l'ordre de grandeur de la largeur d'un arbre binaire de recherche. Appelons en effet $\overline{Z}_n$ la largeur de T_n ; c'est le maximum sur tous les niveaux du nombre de noeuds à chaque niveau :

$$\overline{Z}_n := \max_{k \geq 0} \; Z_k(n)$$

Théorème 13 *(largeur d'un arbre binaire de recherche)*

$$\frac{\overline{Z}_n}{n/\sqrt{\pi \log n}} = 1 + \mathcal{O}(\frac{1}{\sqrt{\log n}}) \quad p.s.$$

lorsque $n \to +\infty$.

<u>Exercice</u> (*) Soit E_n la longueur de cheminement externe, c'est-à-dire la somme des longueurs de u pour tous les u noeuds externes de l'arbre T_n. Montrer que

$$E_n = W_n'(1)$$

où W_n est le polynôme de niveau. En déduire le théorème de Régnier (1989) :

Théorème 14 $\frac{1}{n+1}(E_n - \mathbb{E}(E_n))$ *est une* $\mathcal{F}_n$*-martingale qui converge dans* L^2.

References

[1] K. B. Athreya and P. E. Ney (1972) Branching processes. Springer-Verlag.

[2] J. D. Biggins. (1997). How fast does a general branching random walk spread? Classical and modern branching processes. IMA Vol. Math. Appl. **84** 19–39.

[3] B. Chauvin, M. Drmota and J. Jabbour-Hattab. (2001) The profile of binary search trees. to appear in Annals of Applied Probability

[4] L. Devroye. (1987). Branching processes in the analysis of the height of trees. Acta Inf. **24** 277–298.

[5] M. Drmota. (2000) The Variance of the Height of Digital Search Trees, Acta Inf., submitted.

[6] M. Duflo. (1990) Méthodes récursives aléatoires. Masson.

[7] G. R. Grimmett and D. R. Stirzaker (1992) Probability and Random Processes. 2nd ed. Oxford University Press.

[8] P. Hall and C. C. Heyde. (1980) Martingale Limit Theory and Its Application. Academic Press.

[9] J. Jabbour-Hattab. (2000). Martingales and large deviations for the binary search trees. to appear in Random Structures and Algorithms.

[10] H. M. Mahmoud. (1992) Evolution of Random Search Trees. Wiley.

[11] H. Mahmoud, R.T. Smythe, J. Szymanski (1993) On the structure of random plane-oriented recursive trees and their branches. Random Structures Algorithms 4 (1993), no. 2, 151–176.

[12] C. McDiarmid. (1998) Concentration. in Probabilistic methods for algorithmic discrete mathematics. Ed. Habib, McDiarmid, Ramirez, Alfonsin, Reed. Springer Series in Algorithmics and Combinatorics. **16**, 195-248.

[13] J. Neveu. (1972) Martingales à temps discret. Masson.

[14] V.V. Petrov. (1995) Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Oxford Science Publications.

[15] J. Spencer. (1993) Nine lectures on random graphs. in Ecole d'été de Probabilités de Saint-Flour XXI - 1991. Lecture Notes in Mathematics 1541. Springer.

[16] D. Williams. (1991). Probability with martingales. Cambridge University Press.