

Introduction au domaine de la recherche
La théorie des graphes aléatoires
au service des systèmes de recommandation séquentiels

Francisco Calvillo Marmaneu

Mai 2023

Table des matières

1	Préambule	2
2	Des recommandations aux forêts aléatoires soustractives	2
2.1	Construction du modèle de base	2
2.2	Phénomène de survie et extinction des arbres de la forêt	4
3	Un système de recommandation à deux choix	4
3.1	Deux choix, trois régimes	6
3.1.1	Comportement du système lorsque Z est à queue légère :	6
3.1.2	Comportement du système lorsque Z est à queue lourde :	7
3.1.2.1	Cas $\mathbf{E} \min(Z, W) < \infty$:	7
3.1.2.2	Cas $\mathbf{E} \min(Z, W) = \infty$	8
3.2	Bilan de consistance	9

1 Préambule

De nos jours, la grande majorité des algorithmes de recommandation que nous rencontrons sur les différents sites internet parviennent, d'une façon ou d'une autre, à prendre en compte les intérêts passés de leurs utilisateurs. Parmi ceux-ci, les plus performants sont basés sur des intelligences artificielles. Malgré leur indéniable efficacité, leur fonctionnement peut parfois rester très opaque et nécessitent d'une modélisation mathématique pour une meilleure compréhension de ses mécanismes. Broutin, Lugosi, Devroye et Imbuzeiro Oliveira [1] se sont intéressés à un type système de recommandation qui agit de manière séquentielle et qui s'appuie essentiellement sur l'historique des préférences de chaque individu d'une manière assez simple : à chaque étape le système sélectionne une durée temporelle aléatoire, la soustrait à la date actuelle, puis recommande à l'utilisateur le même sujet qui a été précédemment recommandé à cet instant donné. Dans la section 2, nous verrons que la modélisation d'un tel système conduit de manière naturelle à l'étude des propriétés d'une famille de forêts aléatoires qu'ils appellent *subtractive random forest* (SuRF). L'étude de ces forêts permet de décrire le comportement à long terme du système de recommandation.

Une manière de rendre ce modèle plus réaliste c'est de considérer un système dans lequel l'utilisateur, à chaque étape, au lieu de recevoir une seule recommandation, doit choisir entre deux sujets qui lui sont proposés à l'aide du mécanisme décrit ci-dessus. Ceci suppose une complexification considérable par rapport au système de recommandation à un choix, rendant plus difficile d'avoir une intuition sur son comportement à long terme. La section 3 a pour objectif de définir un tel modèle, de donner des critères de consistance à long terme, et de présenter les premiers résultats obtenus lors de mon stage de pré-thèse, réalisé sous la direction de Gabor Lugosi et en collaboration avec Luc Devroye.

2 Des recommandations aux forêts aléatoires soustractives

2.1 Construction du modèle de base

Nous supposons qu'au départ il existe un ensemble infini de sujets qui peuvent être recommandés, et que les durées aléatoires sélectionnées par le système sont indépendantes et identiquement distribuées. Afin d'établir un modèle mathématique, nous allons commencer par définir un coloriage aléatoire des entiers positifs de la manière suivante. Soit $(Z_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées, prenant des valeurs dans l'ensemble des entiers strictement positifs $\mathbf{N} = \{1, 2, \dots\}$. Dans la suite, nous utiliserons les notations

$$q_i = \mathbf{P}(Z = i) \quad \text{et} \quad p_i = \mathbf{P}(Z \geq i) \quad \text{pour tout } i \geq 1,$$

où Z est une variable aléatoire de même distribution que les Z_i . Pour tout entier négatif $i \in \mathbf{Z}_{\leq 0} = \{0, -1, -2, \dots\}$, posons $C_i = i$. Nous pouvons alors attribuer des couleurs aux entiers positifs $n \in \mathbf{N} = \{1, 2, \dots\}$ par la récursion

$$C_n = C_{n-Z_n}.$$

Chaque couleur correspond alors à un sujet, et s'identifie à un entier de $\mathbf{Z}_{\leq 0}$. Le sujet recommandé par le système à l'instant $n \geq 1$ est donné par sa couleur C_n .

Remarque 1. *Il est intéressant de remarquer que le problème de détermination de la valeur C_n d'un processus au temps n en regardant un nombre aléatoire de pas dans le passé apparaît également dans certains des modèles d'urne et des modèles de marche aléatoire renforcée (voir Pemantle [5]).*

Rappelons que dans le modèle traditionnel de l'urne de Polya, la valeur de C_n représente la couleur de la n -ième boule ajoutée à l'urne, et la loi conditionnelle de C_n , étant donné le passé, est celle d'un échantillon uniforme dans $\{C_1, C_2, \dots, C_{n-1}\}$. Notre C_n diffère du processus de l'urne de Polya dans le fait que le passé est infini et que l'échantillon uniforme est remplacé par un échantillon ayant une distance temporelle par rapport au présent donné par la loi de Z .

Ce processus permet de définir une forêt aléatoire G_1 dont l'ensemble des sommets est l'ensemble des nombres entiers \mathbf{Z} , et dans laquelle deux entiers $n < m$ sont reliés par une arête si et seulement si $m \geq 1$ et $m - Z_m = n$. Ainsi, chaque entier négatif $i \in \mathbf{Z}_{\leq 0}$ se trouve à la racine d'un arbre de la forêt, et l'arbre enraciné en i contient exactement tous les entiers $n \in \mathbf{Z}$ tels que $C_n = i$. Dans cette section, nous utiliserons la notation $T_n^{(i)}$ pour faire référence à l'arbre enraciné en $i \in \mathbf{Z}_{\leq 0}$ au temps $n \geq 1$, et $T^{(i)}$ pour désigner l'arbre à la limite. Autrement dit,

$$T_n^{(i)} = \{1 \leq k \leq n : C_k = i\} \text{ et } T^{(i)} = \bigcup_{n \geq 1} T_n^{(i)}.$$

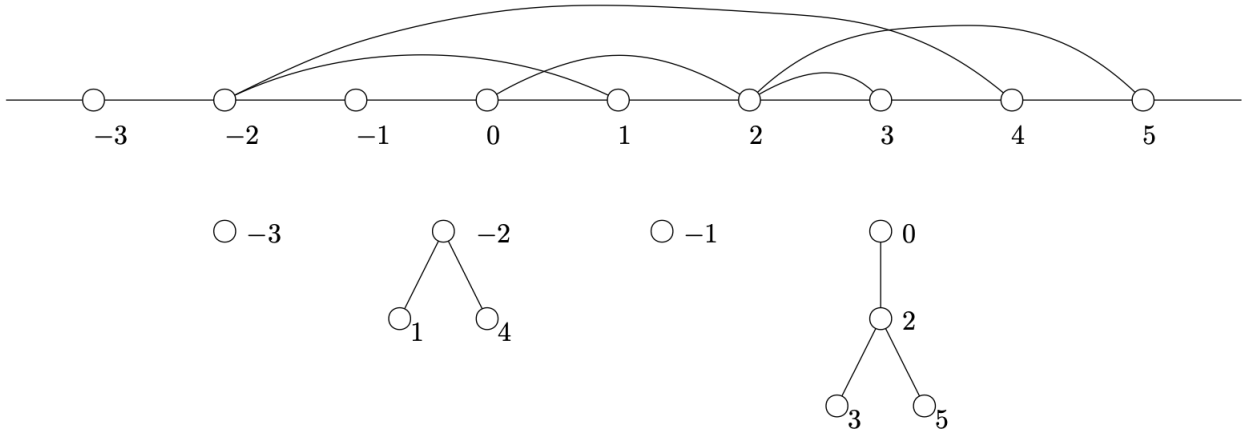


FIGURE 1 – Exemple du processus d'attachement arrêté au temps $n = 5$ et de sa forêt correspondante. Ici, $Z_1 = 3, Z_2 = 2, Z_3 = 1, Z_4 = 6$ et $Z_5 = 3$.

Remarque 2. *Une version très similaire de ce processus aléatoire avait déjà été introduite dans un contexte différent par Hammond et Sheffield [3], qui ont utilisé le coloriage pour définir une marche aléatoire qui converge en échelle vers ce qu'on appelle un mouvement brownien fractionnaire lorsque la queue de la loi de Z est suffisamment lourde.*

2.2 Phénomène de survie et extinction des arbres de la forêt

Ce paragraphe est consacré à la question de savoir si les arbres $T^{(i)}$ de la forêt aléatoire sont finis ou infinis. Les deux théorèmes suivants, prouvés dans [1], mettent en évidence un fort contraste dans le comportement à la limite de la forêt G_1 en fonction de la queue de la variable aléatoire Z . Lorsque Z a une queue légère, c'est-à-dire lorsque $EZ < \infty$, alors un seul arbre survit et le nombre total de sommets d'indice positif de tous les arbres restants est fini, presque sûrement. Cela contraste fortement avec ce qui se passe lorsque Z a une queue lourde : si $EZ = \infty$, alors chaque arbre de la forêt est fini.

Théorème 1. *Supposons que $EZ < \infty$. Alors, presque sûrement, il existe un unique entier $i \leq 0$ tel que $|T^{(i)}| = \infty$.*

Si, de plus, on suppose que $q_1 > 0$, il existe alors une variable aléatoire N et un indice aléatoire $I \in \{0, -1, -2, \dots\}$ tels que, presque sûrement,

$$N < \infty \text{ et } \forall n \geq N, C_n = I .$$

N peut être choisi de sorte à ce qu'il soit un ancêtre commun de tous les sommets $n \geq N$ dans l'arbre $T^{(I)}$.

Théorème 2. *Si $EZ = \infty$ et $q_i > 0$ pour tout $i \geq 1$, alors*

$$\mathbf{P} \left(\exists i \leq 0 : |T^{(i)}| = \infty \right) = 0 .$$

En revenant aux recommandations, cela signifie que si EZ est fini alors à partir d'un certain instant (aléatoire) le système ne recommande qu'un seul sujet ; sinon, tous les sujets ne seront recommandés qu'un nombre fini de fois.

3 Un système de recommandation à deux choix

Dans cette section nous allons nous intéresser à un système qui utilise le même principe pour proposer à chaque fois deux recommandations indépendantes. Nous supposons que l'utilisateur choisira le sujet qu'il préfère parmi ces deux options. Pour construire le modèle mathématique, nous considérons trois suites indépendantes de variables aléatoires

$$\{Z_i, i \geq 1\} , \{W_i, i \geq 1\} \text{ et } \{U_i, i \leq 0\} .$$

Les Z_i, W_i sont supposées indépendantes et identiquement distribués, et les U_i sont à valeurs dans $[0, 1]$. De même que précédemment, nous posons $C_i = i$ pour tous les entiers négatifs $i \leq 0$. Pour tout $n \geq 1$, nous définissons les quantités C_n et U_n par la récursion

$$\begin{cases} C_n = C_{n-Z_n} & \text{et } U_n = U_{n-Z_n} \text{ si } U_{n-Z_n} < U_{n-W_n} \\ C_n = C_{n-W_n} & \text{et } U_n = U_{n-W_n} \text{ sinon} \end{cases} ,$$

de sorte que le sujet regardé par l'utilisateur au temps n soit donné par C_n . La quantité U_n traduit en quelque sorte à quel point l'utilisateur est intéressé par le sujet C_n : au temps $n \geq 1$, le système recommande les sujets C_{n-Z_n} et C_{n-W_n} , et l'utilisateur choisit celui ayant la valeur U plus petite.

Dans la suite, nous considérons le graphe G_2 ayant pour sommets l'ensemble des nombres entiers et dans lequel deux entiers $n < m$ sont reliés par une arête si et seulement si $m \geq 1$ et $m - Z_m = n$ ou $m - W_m = n$.

Remarque 3. Le graphe G_2 n'est rien d'autre que la superposition de deux copies indépendantes de la forêt G_1 étudiée dans la section précédente (l'une construite à l'aide des variables $(Z_i)_{i \geq 1}$, l'autre à l'aide des variables $(W_i)_{i \geq 1}$).

Dans ce nouveau contexte nous ne nous retrouvons plus dans le cadre évoqué dans la Remarque 1, et l'étude directe de la couleur C_n devient techniquement très difficile. L'idée pour gérer ces nouvelles difficultés est de porter notre regard sur le sous-graphe des sommets *atteints* par le sommet n à travers les différentes recommandations successives. Plus proprement, nous définissons par récurrence l'ensemble T_n des sommets atteints par le sommet $n > 0$ en posant $T_n = \emptyset$ pour tout $n \leq 0$ et

$$T_n = \{n, n - Z_n, n - W_n\} \cup T_{n-Z_n} \cup T_{n-W_n}$$

pour $n > 0$. Notons $\mathcal{L}_n = T_n \cap \mathbf{Z}_{\leq 0}$ l'ensemble des sommets dans T_n ayant un indice négatif. \mathcal{L}_n n'est rien d'autre que l'ensemble des sujets qui ont été recommandés dans les différentes étapes qui interviennent dans les chemins temporels partant de n . En particulier, nous avons

$$C_n = \arg \min_{i \in \mathcal{L}_n} U_i \text{ et } U_n = \min_{i \in \mathcal{L}_n} U_i .$$

Même si T_n n'est pas forcément un arbre, nous dirons que les sommets dans \mathcal{L}_n sont des feuilles de T_n .

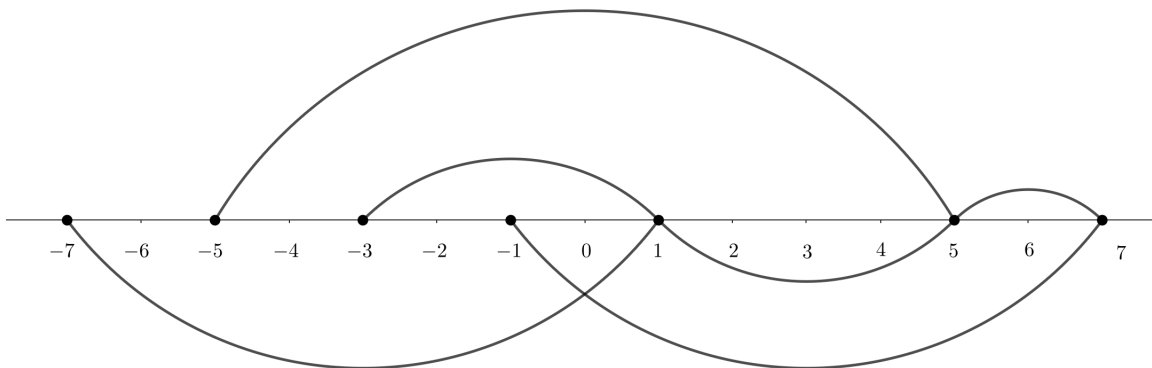


FIGURE 2 – Représentation de T_7 dans le cas où $Z_1 = 4, W_1 = 6, Z_5 = 10, W_5 = 4, Z_7 = 2$ et $W_7 = 8$. Ici, $T_7 = \{7, 5, 1, -1, -3, -5, -7\}$ et $\mathcal{L}_7 = \{-1, -3, -5, -7\}$

Un critère pour évaluer la consistance à long terme d'un tel système de recommandation est de voir quand est-ce que U_n se rapproche de $\min_{i \leq 0} U_i$ lorsque $n \rightarrow \infty$: cela signifie que l'utilisateur est intéressé par les sujets recommandés. Pour identifier les cas dans lesquels le système est consistant, nous avons considéré trois configurations initiales différentes : lorsque les $(U_i)_{i \leq 0}$ sont indépendants et uniformément distribués dans $[0, 1]$, lorsqu'ils sont déterministes et croissants, et lorsqu'ils sont déterministes et décroissants. Nous portons une attention particulière aux variables aléatoires suivantes :

$L_n = |\mathcal{L}_n|$ (le nombre de feuilles atteintes par n)

$M_n = \min \{i : i \in \mathcal{L}_n\}$ (la feuille la plus à gauche dans \mathcal{L}_n)

$R_n = \max \{i : i \in \mathcal{L}_n\}$ (la feuille la plus à droite dans \mathcal{L}_n).

Dans le cas où les $(U_i)_{i \leq 0}$ sont i.i.d. uniformément distribués dans $[0, 1]$, le système est consistant lorsque $L_n \rightarrow \infty$. Si les $(U_i)_{i \leq 0}$ sont croissants, nous avons consistance lorsque $M_n \rightarrow -\infty$. Pour les $(U_i)_{i \leq 0}$ décroissants, le critère de consistance doit être relaxé car en général nous n'aurons pas $R_n \rightarrow 0$: on se contentera tout simplement d'avoir R_n borné.

3.1 Deux choix, trois régimes

Le but de cette partie est d'identifier les cas dans lesquels le système est consistant. Pour la suite, on se donne deux variables aléatoires indépendantes Z et W de même loi que les Z_i . Tout comme pour le modèle à un seul choix, le système présentera différents comportements à long terme en fonction de la queue de la loi de Z . Cependant, nous devons ici distinguer trois cas différents :

- Si Z a une queue légère, de sorte que $\mathbf{E}Z = \sum_{i \geq 1} p_i < \infty$, le système n'a pas les moyens d'explorer suffisamment en arrière dans le temps, et il ne sera consistant que lorsque les $(U_i)_{i \leq 0}$ forment une suite décroissante.
- Si Z a une queue *moyennement lourde*, dans le sens où $\mathbf{E}Z = \sum_{i \geq 1} p_i = \infty$ mais $\mathbf{E} \min(Z, W) = \sum_{i \geq 1} p_i^2 < \infty$, alors le système parvient à explorer le passé de manière suffisamment équilibrée pour garantir sa consistance à long terme dans les trois configurations initiales considérées.
- Si la queue de Z est très lourde, de sorte que $\mathbf{E} \min(Z, W) = \infty$, le système parvient facilement à explorer le passé lointain, mais les sauts temporels sont trop importants pour le faire de manière exhaustive. La consistance à long terme n'est garantie que lorsque les $(U_i)_{i \leq 0}$ forment une suite croissante.

3.1.1 Comportement du système lorsque Z est à queue légère :

Nous considérons d'abord le cas où $\mathbf{E}Z$ est fini. Le théorème suivant nous dit que les trois suites de variables aléatoires sont presque sûrement bornées.

Théorème 3. *Soit $M_\infty = \inf_{n \in \mathbb{N}} M_n$. Si $\mathbf{E}Z < \infty$, alors*

$$M_\infty \text{ est fini presque sûrement.}$$

Ainsi, les suites $(R_n)_{n \in \mathbb{N}}$, $(M_n)_{n \in \mathbb{N}}$ et $(L_n)_{n \in \mathbb{N}}$ sont presque sûrement bornées.

Démonstration. Remarquons que $\mathbf{P}(M_\infty \leq -x) \leq 2 \sum_{n=1}^{\infty} \mathbf{P}(n - Z_n \leq -x) = 2 \sum_{n=1}^{\infty} p_{n+x}$. Comme $\sum_{n \geq 1} p_n = \mathbf{E}Z < \infty$, nous savons que $\sum_{n=1}^{\infty} p_{n+x}$ est fini et tend vers 0 lorsque $x \rightarrow +\infty$. Ainsi, par continuité de la mesure,

$$\mathbf{P}(M_\infty = -\infty) = \lim_{x \rightarrow +\infty} \mathbf{P}(M_\infty \leq -x) = 0.$$

Comme $|M_n| \geq L_n$ et $|M_n| \geq |R_n|$, ceci entraîne que les suites $(L_n)_{n \in \mathbb{N}}$ et $(R_n)_{n \in \mathbb{N}}$ sont aussi presque sûrement bornées. \square

Remarque 4. *Le modèle à un choix nous permet de justifier ce dernier résultat d'une manière plus visuelle. En effet, d'après le Théorème 1, un seul arbre de la forêt G_1 survit. Comme G_2 s'obtient en superposant deux copies indépendantes de G_1 , cela nous dit qu'il y aura au plus deux sujets (et au moins un) qui pourront être atteints par un nombre infini de sommets, presque sûrement.*

Observons que le Théorème 3 implique que les suites (M_n) et (L_n) ne peuvent pas tendre vers l'infini, ni presque sûrement (ce qui est clair) ni en probabilité.

3.1.2 Comportement du système lorsque Z est à queue lourde :

Lorsque $EZ = \infty$, les sauts temporels réalisés par le système sont suffisamment grands pour qu'il puisse aller loin dans le temps. Avant de distinguer les cas selon la finitude de $E \min(Z, W)$, commençons par énoncer le résultat suivant, qui montre que l'hypothèse sur EZ suffit pour prouver que la suite M_n tend vers $-\infty$ presque sûrement.

Proposition 1. *Si $q_i > 0$ pour tout $i \in \mathbb{N}$ et $EZ = \infty$, alors*

$$M_n \rightarrow -\infty \text{ presque sûrement.}$$

Démonstration. Ce résultat découle du comportement du modèle lorsqu'il n'y a qu'une seule recommandation. Notons T_n^Z le sous-ensemble de T_n des sommets qui sont atteints par le sommet n uniquement à l'aide des arêtes données par les variables Z_i . Plus précisément, nous définissons T_n^Z par récurrence en posant $T_n^Z = \emptyset$ pour tout $n \leq 0$, et

$$T_n^Z = \{n, n - Z_n\} \cup T_{n-Z_n}^Z$$

pour tout $n \geq 1$. D'après le Théorème 2 nous savons que, presque sûrement, pour tout $k \leq 0$, il existe au plus un nombre fini d'entiers $n \geq 1$ tels que $k \in T_n^Z$. Ainsi, la suite $(\min T_n^Z)_{n \in \mathbb{N}}$, qui appartient à $\{0, -1, -2, \dots\}$, ne peut pas prendre la même valeur un nombre infini de fois, donc elle diverge vers $-\infty$ lorsque $n \rightarrow \infty$. Ceci conclut, puisque $M_n \leq \min T_n^Z$ pour tout $n \geq 1$. □

3.1.2.1 Cas $E \min(Z, W) < \infty$:

Si $E \min(Z, W)$ est fini, les pas temporels sont suffisamment petits pour que T_n contienne un grand nombre de points. Nous allons voir par la suite que lorsque $EZ = \infty$ et $E \min(Z, W) < \infty$, le nombre de feuilles atteintes par le sommet n tend vers l'infini et que (R_n) reste bornée presque sûrement. L'idée clé est d'introduire l'ensemble S_n des sommets atteints par n en utilisant uniquement les sauts les plus courts. Plus formellement, nous définissons S_n par récurrence en posant $S_n = \emptyset$ pour tout $n \leq 0$, et

$$S_n = \{n, n - \min(Z_n, W_n)\} \cup S_{n - \min(Z_n, W_n)}$$

pour tout $n \geq 1$. Observons que le graphe dont l'ensemble des sommets est \mathbf{Z} et dont l'ensemble des arêtes est $\{\{n, n - \min(Z_n, W_n)\}, n \geq 1\}$ n'est rien d'autre qu'une forêt aléatoire soustractive construite avec la suite de variables aléatoires $(\min(Z_i, W_i))_{i \geq 1}$. Sous les hypothèses $E \min(Z, W) < \infty$ et $q_1 > 0$, le Théorème 1 nous dit qu'il existe une variable aléatoire N , finie presque sûrement, et telle que pour tout $n \geq N$, $N \in S_n$.

Théorème 4. *Supposons que $\mathbf{E} \min(Z, W) < \infty$ et que $q_1 > 0$. Alors la suite $(R_n)_{n \in \mathbf{N}}$ est bornée presque sûrement.*

Démonstration. Soit N une variable aléatoire comme ci-dessus. Pour tout $n \geq N$, nous savons que $N \in S_n$. En particulier, ceci implique que $T_N \subset T_n$, et donc que $|R_N| \geq |R_n|$. Ainsi,

$$\sup_{n \in \mathbf{N}} |R_n| \leq \max_{1 \leq k \leq N} |R_k| < \infty .$$

□

Théorème 5. *Supposons que $q_1 > 0$, que $\mathbf{E} \min(Z, W) < \infty$ et que $\mathbf{E}Z = \infty$. Alors*

$$L_n \rightarrow \infty \text{ presque sûrement.}$$

Idée de preuve. Cette preuve étant un peu plus technique, nous n'entrerons pas dans les détails. L'idée est d'utiliser la chaîne S_n et le Théorème 1 pour justifier le fait que T_n est grand. De chaque sommet dans S_n sort une arête donnée par la loi temporelle $\max(Z, W)$, qui a des fortes chances de tomber sur un sommet d'indice négatif. □

3.1.2.2 Cas $\mathbf{E} \min(Z, W) = \infty$

Si la queue de $\min(Z, W)$ est lourde, les sauts temporels deviennent trop longs et le système ne sera pas consistant au sens presque sûr lorsque les (U_i) sont uniformément distribués dans $[0, 1]$, ou lorsqu'ils sont rangés par ordre décroissant.

Théorème 6. *Supposons que $\mathbf{E} \min(Z, W) = \infty$. Alors, presque sûrement,*

- (i) $L_n \rightarrow \infty$ lorsque $n \rightarrow \infty$,
- (ii) la suite $(R_n)_{n \in \mathbf{N}}$ n'est pas bornée.

Démonstration. Observons d'abord que si $\min(Z_n, W_n) \geq n$, alors $L_n \in \{1, 2\}$. Ainsi, pour prouver (i) il suffit de prouver que, presque sûrement, l'événement $\min(Z_n, W_n) > n$ a lieu infiniment souvent. Ceci découle du lemme Borel-Cantelli, puisque les événements $(\{\min(Z_n, W_n) \geq n\})_{n \in \mathbf{N}}$ sont indépendants, et

$$\sum_{n \geq 1} \mathbf{P}(\min(Z_n, W_n) \geq n) = \mathbf{E} \min(Z, W) = \infty .$$

De même, les événements $(\{\min(Z_n, W_n) \geq 2n\})_{n \in \mathbf{N}}$ sont indépendants et

$$\sum_{n \geq 1} \mathbf{P}(\min(Z_n, W_n) \geq 2n) = \sum_{n \geq 1} p_{2n}^2 = \infty .$$

Ainsi, d'après le lemme de Borel-Cantelli, nous avons $\min(Z_n, W_n) \geq 2n$ infiniment souvent, presque sûrement, et donc que $R_n \leq -n$ infiniment souvent, ce qui prouve (ii). □

Cependant, le dernier résultat que nous allons énoncer nous dit que, sous une hypothèse supplémentaire, le système est faiblement consistant lorsque les U_i sont uniformément distribués dans $[0, 1]$. Notons $r_n = \mathbf{P}(0 \in T_n^Z)$, où T_n^Z est l'ensemble défini dans la preuve de la Proposition 1.

Théorème 7. *Si $\mathbf{E} \min(Z, W) = \infty$ et $\sum_{n \geq 0} r_n^2 < \infty$, alors*

$$L_n \rightarrow \infty \text{ en probabilité.}$$

3.2 Bilan de consistance

	$\mathbf{E}Z < \infty$	$\mathbf{E}Z = \infty$ et $\mathbf{E} \min(Z, W) < \infty$	$\mathbf{E} \min(Z, W) = \infty$
	comportement p.s.		
$M_n \rightarrow -\infty$ p.s.	Non (Th. 3)	Oui (Prop. 1)	Oui (Prop. 1)
$L_n \rightarrow \infty$ p.s.	Non (Th. 3)	Oui (Th. 5)	Non (Th. 6)
R_n bornée p.s.	Oui (Th. 3)	Oui (Th.4)	Non (Th. 6)
	comportement faible		
$M_n \xrightarrow{\mathbf{P}} -\infty$	Non (Th. 3)	Oui (Prop. 1)	Oui (Prop. 1)
$L_n \xrightarrow{\mathbf{P}} \infty$	Non (Th. 3)	Oui (Th. 5)	Oui si $\sum r_n^2 < \infty$ (Th. 7)
R_n est tendue	Oui (Th. 3)	Oui (Prop. 4)	?

Références

- [1] Nicolas BROUTIN et al. *Subtractive random forests*. 2022. DOI : 10.48550/ARXIV.2210.10544. URL : <https://arxiv.org/abs/2210.10544>.
- [2] P ERDÖS, W FELLER et H POLLARD. “A theorem on power series”. In : *Bull. Amr. Math. Soc* 55 (1949), p. 201-204.
- [3] Alan HAMMOND et Scott SHEFFIELD. “Power law Pólya’s urn and fractional Brownian motion”. In : *Probability Theory and Related Fields* 157.3 (2013), p. 691-719. DOI : 10.1007/s00440-012-0468-6. URL : <https://doi.org/10.1007/s00440-012-0468-6>.
- [4] Jan Lukas IGELBRINK et Anton WAKOLBINGER. “Asymptotic Gaussianity via coalescence probabilities in the Hammond-Sheffield urn”. In : *Latin American Journal of Probability and Mathematical Statistics* 20.1 (2023), p. 53. DOI : 10.30757/alea.v20-04. URL : <https://doi.org/10.30757%2Falea.v20-04>.
- [5] Robin PEMANTLE. “A survey of random processes with reinforcement”. In : *Probability Surveys* 4.none (2007), p. 1-79. DOI : 10.1214/07-PS094. URL : <https://doi.org/10.1214/07-PS094>.