HAL
open science

Le modèle probabiliste de l'alignement local de deux séquences

Pierre Nicodème

To cite this version:

Pierre Nicodème. Le modèle probabiliste de l'alignement local de deux séquences. [Rapport de recherche] RR-3201, INRIA. 1997. inria-00073488

HAL Id: inria-00073488

https://inria.hal.science/inria-00073488v1

Submitted on 24 May 2006

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

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

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Le modèle probabiliste de l'alignement local de deux séquences

Pierre Nicodème

N`3201
Juin 1997

THÈME 2

Rapport de recherche

ISSN 0249-6399

Le modèle probabiliste de l'alignement local de deux séquences

Pierre Nicodème

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n° 3201 — Juin 1997 — 43 pages

Résumé : Nous décrivons le modèle probabiliste utilisé pour calibrer l'alignement local de deux séquences, quand un schéma de score additif est utilisé. Ce modèle est utilisé en particulier par le logiciel BLAST (Basic Local Alignment Search Tool), qui est employé fréquemment par les biologistes. Asymptotiquement, la probabilité associée au score d'un alignement optimal vérifie une loi des valeurs extrêmes. Nous analysons différents aspects des outils mathématiques (modèles de marche aléatoire, décomposition de Wiener-Hopf, théorie du renouvellement) utilisés dans ce modèle probabiliste.

(Abstract: pto)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

The probability model of local alignment of two sequences

Abstract: We describe the probability model of local alignment of two sequences, when an additive scoring scheme is used. This probability model is used in particular by the software BLAST (Basic Local Alignment Search Tool), frequently used by biologists. Asymptotically, the probability associated to the score of an optimal alignment verifies an extreme values distribution. We analyse some aspects of the mathematical tools (random walk models, Wiener-Hopf decomposition, renewal theory) underlying this probability model.

LE MODÈLE PROBABILISTE DE L'ALIGNEMENT LOCAL DE DEUX SÉQUENCES

RÉSUMÉ. Nous décrivons le modèle probabiliste utilisé pour calibrer l'alignement local de deux séquences, quand un schéma de score additif est utilisé. Ce modèle est utilisé en particulier par le logiciel BLAST (Basic Local Alignment Search Tool), qui est employé fréquemment par les biologistes. Asymptotiquement, la probabilité associée au score d'un alignement optimal vérifie une loi des valeurs extrêmes. Nous analysons différents aspects des outils mathématiques (modèles de marche aléatoire, décomposition de Wiener-Hopf, théorie du renouvellement) utilisés dans ce modèle probabiliste.

1. INTRODUCTION

Nous allons considérer le modèle mathématique sous-jacent à l'alignement local de deux séquences, dans le cas où les alignements $1-1$ considérés sont sans trous, et où l'on utilise un schéma de scores additif. Étant donné deux séquences $s_1 \dots s_{n_1}$ et $t_1 \dots t_{n_2}$ sur un même alphabet Σ, un alignement local $\mathcal{A}$ de longueur l est une suite de paires

$$\{(s_{j_1}, t_{j_2}), (s_{j_1+1}, t_{j_2+1}), \dots, (s_{j_1+l}, t_{j_2+l})\},$$

avec $1 \leq j_1 \leq n_1$, $1 \leq j_2 \leq n_2$, $j_1 + l \leq n_1$, et $j_2 + l \leq n_2$. Étant donné d'autre part une matrice $\mathcal{M}$ associant à toute paire $\{a_i, a_j\} \in \Sigma \times \Sigma$ un score, le score de l'alignement est égal à la somme du score des paires qui le constituent.

Dans les cas d'alignement de séquences biologiques qui nous intéressent, et plus particulièrement de séquences protéiques, nous considérerons des matrices de similarité entre acides aminés qui associent des scores positifs à des paires d'acides aminés de caractéristiques physico-chimiques voisines, et des scores négatifs à celles qui ont des caractéristiques différentes. Nous nous intéresserons alors aux alignements de scores élevés et tout particulièrement à l'alignement optimal, qui est l'alignement de meilleur score. En négligeant les effets de bord, l'espace de recherche a comme dimension le produit $n_1 \times n_2$ de la taille des deux séquences.

En considérant d'autre part une distribution de probabilité sur l'alphabet Σ, et en considérant que les tirages successifs de lettres sont indépendants, nous en déduisons immédiatement une distribution de probabilité sur l'alphabet $\Sigma \times \Sigma$ sur lequel nous avons défini les scores. Nous avons ainsi défini une variable aléatoire X de $\Sigma \times \Sigma$ sur $\mathbb{R}$, avec

$$X(a_1, a_2) = \mathcal{M}(a_1, a_2), \quad \{(a_1, a_2) \in \Sigma \times \Sigma\},$$

Nous sommes donc amenés à étudier le problème général suivant.

Énoncé probabiliste associé à une recherche d'alignement optimal. Soit X_i une suite de variables aléatoires i.i.d, de distribution X. Soit S_k les sommes partielles, $S_k = \sum_{i=1}^{k} X_i$. On cherche à connaître la distribution de $M(n) = \max_{1 \leq j \leq k \leq n}(S_k - S_j)$.

La variable $M(n)$ correspond au score de l'alignement optimal de deux séquences de longueur n_1 et n_2 telles que $n_1 \times n_2 = n$.

Nous démontrerons que

$$Pr\left(M(n) > \frac{log(n)}{\lambda} + x\right) = 1 - e^{-Ke^{-\lambda x}}. \tag{1}$$

où λ peut être calculé comme racine d'une équation, et K possède un développement en série de convergence géométrique. Asymptotiquement, quand $n \to \infty$, cette recherche n'a de sens que si $E(X) < 0$; dans le cas contraire, $M(n) \to \infty$ avec probabilité 1 par la loi des grands nombres.

Nous allons considérer la marche aléatoire définie dans le demi-plan d'abscisses positives par la suite des points de coordonnées $\{i, S_i\}$, $i \in \mathbb{N}$, que nous noterons par la suite (S_i). Nous rechercherons alors la section $S_{[k,l]} = S_l - S_k$ de valeur maximale. La démonstration repose principalement sur des propriétés de la marche aléatoire (S_i). Elle utilise également des résultats de la théorie du renouvellement (renewal theory). La première démonstration de l'équation 1, représentative d'une loi des valeurs extrêmes, et des formules donnant les paramètres K et λ, dans le cas où X est une variable continue, est due à Iglehart (1972). Cette démonstration a été généralisée par Karlin et Dembo au cas discret (1992) et appliquée par Altschul au cas de la recherche de similarités entre séquences biologiques (protéines ou ADN) dans le logiciel d'alignement de séquences BLAST [1].

Nous ferons fréquemment référence aux articles de Iglehart [13], et de Karlin et Dembo [14], ainsi qu'aux livres de Chung [3] et Durrett [8] et au cours de DEA de Philippe Robert sur les Files d'Attente (Paris VI, Laboratoire de Probabilités, 1993-1994). (On trouvera également de nombreux résultats relatifs aux marches aléatoires dans Gut [12]). Nous avons approfondi certaines démonstrations, et nous avons donné la preuve d'un résultat concernant le paramètre K qui était annoncé par Karlin sans démonstration.

Nous donnons à présent le schéma général de la démonstration, que nous expliciterons de manière plus formelle au prochain paragraphe. (On pourra se reporter au tableau 1 pour les notations).

La marche aléatoire ($S_i = \sum_{j=1}^{i} X_j$) est décomposée en sections en utilisant des *points d'échelle* négatifs; ces points correspondent à des temps de record négatif successifs. La figure 1 donne un exemple de cette décomposition.

Si q_i est le ième point d'échelle, la marche aléatoire y a la valeur S_{q_i}. On définit récursivement le point d'échelle q_{i+1} par $\{q_{i+1} = \min(y;\ \forall x,\ q_i \leq x < y;\ S_y < S_x), q_0 = 0\}$.

Les sections de la marche aléatoire $S_{[q_i,q_{i+1}]}$, situées entre les temps d'échelle, sont alors des répliques probabilistes de la première section $S_{[q_1,q_2]}$, qui est la portion de la marche aléatoire partant de l'origine et s'arrêtant au temps d'arrêt ζ^- de première atteinte du demi-plan négatif. On remarquera que la section $S_{[x,y]}$ de valeur maximale commence certainement en l'un des points d'échelle q_i, et qu'elle est même le record Q_i pour la section commençant à ce point d'échelle. La variable $M(q_m)$ correspondra au maximum des variables Q_i, $i \in [1, m]$, des m premières sections, et la remarque précédente montre que tous les Q_i ont même distribution que Q_1.

L'identité de Spitzer [27] donne, pour une marche aléatoire correspondant à la somme d'une variable aléatoire X de moyenne strictement négative $E(X) < 0$, la loi des sommes $Z^+ = S_{\zeta^+}$ (resp. $Z^- = S_{\zeta^-}$) obtenus au temps d'arrêt ζ^+ (resp. ζ^-) du premier passage dans le demi-plan d'abscisse positive (resp. négative).

Partant de ces identités pour la loi de Z^+, en décomposant la marche aléatoire infinie en la partie précédant ζ^+, et le reste de la marche aléatoire, on reliera la queue de distribution du record absolu S_R de la marche aléatoire à la loi de Z^+.

En effectuant ensuite une décomposition "verticale" de la marche aléatoire, on trouvera une décomposition asymptotique faisant intervenir $q_1 = \zeta^-$, et on reliera la queue de distribution de Q_1 à celle de S_R.

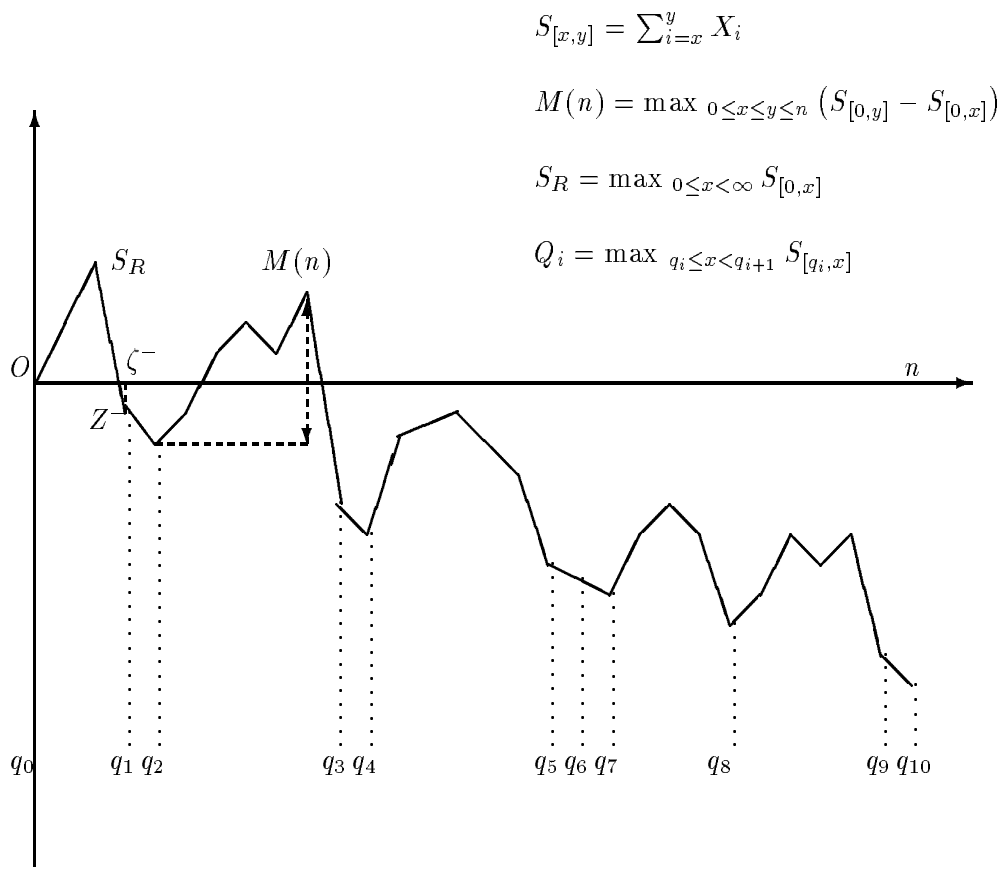

FIGURE 1. Décomposition de la marche aléatoire pour la démonstration de la formule de Karlin

$$\mathcal{L}(Z^+) \Longrightarrow \mathcal{Q}(S_R) \Longrightarrow \mathcal{Q}(Q_1) \Longrightarrow \mathcal{Q}(M(q_i)) \Longrightarrow \mathcal{Q}(M(n)),$$

avec $\mathcal{L}(X)$ et $\mathcal{Q}(X)$ respectivement loi et queue de la distribution de X

FIGURE 2. Schéma de la démonstration d'Iglehart-Karlin:

X	l'incrément de la marche aléatoire, avec $E(X) < 0$ et $P(X > 0) > 0$
$F(y)$	la distribution de X: $F(y) = P(X < y)$
λ	la solution en t de l'équation $E(e^{tX}) = 1$
S_R	la valeur record positive absolue de la marche aléatoire (S_i) $S_R = \max_{i \geq 1} S_{[0,i]}$
t_R	le temps minimum auquel le record absolu S_R est atteint $t_R = \{\min i; S_i = S_R\}$
$R(y)$	la distribution de S_R $R(y) = P(S_R < y)$
$M(n)$	le maximum obtenu par une section de la marche aléatoire dans l'intervalle $[1, n]$ $M(n) = \max S_{[k,j]} = \max_{1 \leq j \leq k \leq n}(S_k - S_j)$.
ζ^+	le temps d'arrêt correspondant au premier passage de la marche aléatoire dans le demi-plan d'ordonnées strictement positives $\zeta^+ = \{\min i, S_i > 0\}$
Z^+	la valeur de la somme partielle pour le temps d'arrêt ζ^+ $Z^+ = S_{\zeta^+}$
ζ^-	le temps d'arrêt correspondant au premier passage de la marche aléatoire dans le demi-plan d'ordonnées négatives $\zeta^- = \{\min i, S_i \leq 0\}$
Z^-	la valeur de la somme partielle pour le temps d'arrêt ζ^- $Z^- = S_{\zeta^-}$
$L_{Z^+}(y)$	la distribution de la variable Z^+ $L_{Z^+}(y) = P(0 < S_{\zeta^+} < y)$
q_i	les temps d'arrêt correspondant aux records négatifs successifs (variables à échelle) $q_0 = 0$, $q_i = \{\min j, S_j < S_{q_{i-1}}\}$
Q_i	le maximum des sommes partielles commençant en q_{i-1} et finissant avant q_i: $Q_i = \{\max S_{[q_i,k]},\ q_i \leq k < q_{i+1}\}$
$Q(y)$	la distribution de la variable Q_1 $Q(y) = P(0 \leq Q_1 < y)$

TABLEAU 1. Notations utilisées.

1.1. Plan de la démonstration.

1. *Calcul de $R(y)$. (Record absolu).* Des résultats de marche aléatoire (identité de Spitzer) donnent la transformée de Laplace de la distribution de Z^+. Cette identité est obtenu en utilisant la factorisation de Wiener-Hopf [3, 9, 23].

   On décompose la marche aléatoire infinie en:
   - une section s'arrêtant en ζ^+;
   - le reste de la marche aléatoire.

   Le maximum de cette deuxième section donne lieu à une convolution, et on obtient pour $R(y)$ une équation de renouvellement dépendant de $L(y)$.

   Le théorème du renouvellement nous donne la solution asymptotique de cette équation. On obtient

$$1 - R(y) \to C_1 e^{-\lambda y}, \text{ quand } y \to \infty.$$

2. *Calcul de $Q(y)$. (Record de la première section).* On relie la distribution de Q_1 à celle de S_R. Ceci se fait en décomposant la marche aléatoire par rapport à un seuil y et en comparant

Q_1 à ce seuil. On obtient

$$Pr[S_R > y] = Pr[Q_1 \geq y] + Pr[S_{q_1} = \eta] Pr[S_{R[q_1 \ldots \infty]} > y - \eta, Q_1 < y],$$

où $S_{R[q_1 \ldots \infty]}$ est le record absolu d'une marche aléatoire commençant en q_1. La deuxième partie de l'équation introduit une intégrale de convolution qu'on étudiera asymptotiquement, quand $y \to \infty$.

On en déduit que

$$1 - Q(y) \to C_2 e^{-\lambda y}.$$

La distribution de la variable Z^- intervient dans le calcul de C_2.

3. *Calcul de la queue de distribution de $M(n)$.* On remarquera que $M(q_m)$, ou variation positive maximum sur $[1 \ldots q_m]$ est le maximum des $Q_j, j \in [1, m]$; en effet toute excursion de variation maximum commence nécessairement à un point d'échelle des records décroissants (voir figure 1).

   On a donc $Pr[M(q_m) < y] = Q^m(y)$. Ceci donne la distribution de $M(n)$ pour les valeurs de n égales aux q_i. L'application à la limite ($i \to \infty$) du théorème de Blackwell permet d'obtenir $M(n)$ pour tout n.

2. Identité de Spitzer et factorisation de Wiener-Hopf

Nous considérons la marche aléatoire associée aux sommes partielles $S_n = \sum_{i=1}^n X_n$ où les variables X_n sont indépendantes et distribuées comme une variable donnée X. Nous allons chercher à obtenir la loi de la variable $Z^- = S_{\zeta^-}$, où ζ^- est le temps d'arrêt d'atteinte du demi-plan négatif par la marche aléatoire.

Cette loi est obtenue grâce à une double décomposition de Wiener-Hopf du quasi-inverse de $E\left(e^{-\xi X}\right)$.

On peut trouver cette démonstration dans Karlin & Taylor [16] qui utilise les fonctions caractéristiques; nous suivrons cependant Neveu [22] et Robert [24] pour la démonstration de la décomposition de Wiener-Hopf que nous ferons au paragraphe 2.2, et Chung [3] pour son application au calcul de la loi de Z^-.

La factorisation de Wiener-Hopf associe à la quasi-inverse (marquée par une variable u, avec $|u| < 1$) de la transformée de Laplace de la distribution de X deux fonctions ϕ_+ et ϕ_-:

$$\frac{1}{1 - uE(e^{-\xi X})} = \phi_+(u, \xi)\phi_-(u, \xi); \tag{2}$$

les fonctions ϕ_+ et ϕ_- sont analytiques et respectivement bornées dans les demi-plans $\Re(\xi) \geq 0$ et $\Re(\xi) \leq 0$; cette factorisation est unique.

La double décomposition mentionnée précédemment est,

- d'une part un développement exp-log du quasi-inverse connu sous le nom de formule de Spitzer.
- et d'autre part une décomposition faisant intervenir les temps d'arrêt ζ^- et ζ^+,

Compte tenu de l'unicité des fonctions ϕ^+ et ϕ^-, cette double décomposition de Wiener-Hopf nous donnera deux identités; nous tirerons de l'une d'elle la loi de Z^-.

2.1. **Décomposition de Wiener-Hopf en exp-log.** Nous appliquons la décomposition de Wiener-Hopf à la marche aléatoire (S_n) en utilisant un développement exp-log du quasi-inverse:

$$\frac{1}{1-x} = \exp\left(\log\left\{\frac{1}{1-x}\right\}\right) = \exp\left(\sum_{i=1}^{\infty} \frac{x^i}{i}\right);$$

nous obtenons:

$$\frac{1}{1-uE(e^{-\xi X})} = \phi_+\phi_- = \exp\left(\sum_{i=1}^{\infty}\frac{u^i}{i}\left[E(e^{-\xi X})\right]^i\right); \tag{3}$$

Les X_i étant indépendants, nous avons $E(e^{-\xi S_n}) = \left[E(e^{-\xi X})\right]^n$, et donc

$$\frac{1}{1-uE(e^{-\xi X})} = \phi_+\phi_- = \exp\left(\sum_{i=1}^{\infty}\frac{u^i}{i}E(e^{-\xi S_i})\right), \tag{4}$$

que nous pouvons réécrire:

$$\frac{1}{1-uE(e^{-\xi X})} = \prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}\right)\right), \tag{5}$$

mais nous avons

$$E(e^{-\xi S_i}) = E(e^{-\xi S_i}; S_i > 0) + E(e^{-\xi S_i}; S_i \leq 0);$$

donc

$$\frac{1}{1-uE(e^{-\xi X})} = \phi_+\phi_- = \prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}; S_i > 0\right)\right)\prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}; S_i \leq 0\right)\right). \tag{6}$$

Dans le premier produit infini, on a $\left|e^{-\xi S_n}\right| \leq 1$, puisque $S_n > 0$ et $\Re\xi \geq 0$; nous en déduisons

$$\left|\prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}; S_i > 0\right)\right)\right| < \prod_{i=1}^{\infty}\exp\left(\frac{|u|^i}{i}\right) = \exp\left(\sum_{i=1}^{\infty}\frac{|u|^i}{i}\right) = \frac{1}{1-|u|},$$

et comme $|u| < 1$, ce produit infini est borné. De la même manière, le produit infini conditionné par $\{S_i \leq 0\}$ est borné pour $\Re\xi \leq 0$. La décomposition de Wiener-Hopf étant unique, on obtient

$$\phi_+ = \prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}; S_i > 0\right)\right) \tag{7}$$

$$\phi_- = \prod_{i=1}^{\infty}\exp\left(\frac{u^i}{i}E\left(e^{-\xi S_i}; S_i \leq 0\right)\right) \tag{8}$$

2.2. **Factorisation de Wiener-Hopf suivant les temps d'arrêt ζ^+ et ζ^-.** Nous allons maintenant donner une deuxième décomposition de Wiener-Hopf et la preuve de l'unicité de la décomposition.

Nous avons besoin au préalable d'un lemme disant que la marche aléatoire prise après un temps d'arrêt est une réplique probabiliste de la marche aléatoire initiale. On trouve ce lemme dans Chung [3], Robert [24] et Durrett [8] (p.177).

Lemme 1. *Soit X_n une suite de variables aléatoires i.i.d. et $\mathcal{F}_n$ la tribu engendrée par $X_1, \ldots, X_n$. Soit α un temps d'arrêt $\Omega \to \{\mathbb{N}\bigcup\infty\}$, relativement à la filtration $(\mathcal{F}_n)_n \geq 0$; conditionnellement à l'événement $\{\alpha < \infty\}$, la suite $(S_{\alpha+n} - S_\alpha)$ est indépendante du couple (S_α, α), et a même loi que la marche aléatoire initiale (S_n)*

Preuve. Soit $p \in \mathbb{N}$ et f, g deux fonctions Boréliennes positives, on a

$$E\left(\ f(S_{n+\alpha} - S_\alpha)g(S_\alpha)1_{\{\alpha=p\}}\right) = E\left(\ f(S_{n+p} - S_p)g(S_p)1_{\{\alpha=p\}}\right);$$

En remarquant que $g(S_p)1_{\{\alpha=p\}}$ est $\mathcal{F}_n$ mesurable et que $f(S_{n+p} - S_p)$ ne dépend que des $(X_k)_{k\geq p+1}$, l'indépendance des (X_n) nous donne

$$\begin{aligned} E\left(f((S_{n+\alpha} - S_\alpha)g(S_\alpha)1_{\{\alpha=p\}}\right) &= E\left(f(S_{n+p} - S_p)\right) E\left(g(S_p)1_{\{\alpha=p\}}\right) \\ &= E\left(f(S_n)\right) E\left(g(S_\alpha)1_{\{\alpha=p\}}\right), \end{aligned}$$

et donc $(S_{\alpha+n} - S_\alpha)$ a même loi que S_n et est indépendant du couple (S_α, α). □

Le théorème donnant la décomposition de Wiener-Hopf est alors:

Théorème 2. *Soit X une variable aléatoire. Pour tout u, tel que $|u| < 1$, il existe deux fonctions $\phi_+(u,.)$, $\phi_-(u,.)$ vérifiant*

1.

$$\frac{1}{1 - uE(e^{-\xi X})} = \phi_+(u,\xi)\phi_-(u,\xi). \tag{9}$$

2. *La fonction $\phi_-(u,.)$ est bornée ainsi que son inverse dans le demi-plan $\{\Re\xi \leq 0\}$; la fonction $\phi_+(u,.)$ est bornée ainsi que son inverse dans le demi-plan $\{\Re\xi \geq 0\}$ et*

$$\lim_{\Re\xi\to\infty} \phi_+(u,\xi) = 1.$$

Une telle décomposition est unique.

Nous considérons un temps d'arrêt α (tel que ζ^+ ou ζ^-), bien défini sur la filtration $\mathcal{F}_n$ construite sur la suite X_i. Nous allons décomposer la marche aléatoire à partir de ce temps d'arrêt: Considérons la série génératrice associée à la transformée de Laplace de la distribution S_n; nous avons

$$E\left(\sum_{n=0}^{\infty} u^n e^{-\xi S_n}\right) = E\left(\sum_{n=0}^{\infty} \left[ue^{-\xi X}\right]^n\right) = \frac{1}{1 - uE(e^{-\xi X})}. \tag{10}$$

Nous décomposons le premier terme de l'équation 10 en

$$E\left(\sum_{n=0}^{\infty} u^n e^{-\xi S_n}\right) = E\left(\sum_{n=0}^{\alpha-1} u^n e^{-\xi S_n}\right) + E\left(\sum_{n=\alpha}^{\infty} u^n e^{-\xi S_n}\right), \tag{11}$$

étant entendu que sur l'ensemble $\{\alpha = \infty\}$ la première somme est $\sum_{n=0}^{\infty}$ et la deuxième est vide et égale à zéro. La dernier terme de l'équation 11 peut être réécrit comme

$$E\left(\sum_{n=\alpha}^{\infty} u^n e^{-\xi S_n}\right) = E\left(\sum_{n=0}^{\infty} u^{\alpha+n} e^{-\xi S_{\alpha+n}}\right) = E\left(u^\alpha e^{-\xi S_\alpha} \sum_{n=0}^{\infty} u^n e^{-\xi(S_{\alpha+n} - S_\alpha)}\right). \tag{12}$$

Le lemme 1 nous dit que $S_{\alpha+n} - S_\alpha$ a la même loi que S_n et est indépendant de S_α et de α. Le membre de droite de 12 est donc égal à

$$E\left(u^\alpha e^{-\xi S_\alpha}\right) E\left(\sum_{n=0}^{\infty} u^n e^{-\xi S_n}\right) = E\left(u^\alpha e^{-\xi S_\alpha}\right) \frac{1}{1 - uE(e^{-\xi X})}.$$

Nous pouvons donc réécrire 10 sous la forme

$$\frac{1}{1 - uE(e^{-\xi X})}\left[1 - E\left(u^\alpha e^{-\xi S_\alpha}\right)\right] = E\left(\sum_{n=0}^{\alpha-1} u^n e^{-\xi S_n}\right). \tag{13}$$

Nous avons donc, en prenant $\alpha = \zeta^-$, temps d'atteinte du demi-plan d'ordonnées négatives,

$$\frac{1}{1 - uE(e^{-\xi X})} = \frac{E\left(\sum_{n=0}^{\zeta^- - 1} u^n e^{-\xi S_n}\right)}{1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right)}. \tag{14}$$

Notons

$$\phi_-(u,\xi) = \frac{1}{1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right)}. \tag{15}$$

Comme $S_{\zeta^-} \leq 0$ et $|u| < 1$, sur $\Re\xi \leq 0$, nous avons

$$E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right) < |u|,$$

et donc

$$1 - |u| \leq |1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right)| \leq 1 + |u|;$$

donc ϕ_- et son inverse sont bornées sur $\{\Re\xi \leq 0\}$.

Notons à présent

$$\phi_+(u,\xi) = E\left(\sum_{0 \leq n < \zeta^-} u^n e^{-\xi S_n}\right). \tag{16}$$

Nous devons montrer que ϕ_+ possède aussi les propriétés énoncées dans le théorème 2.

Nous allons utiliser à cette fin un argument probabiliste en suivant Neveu [22] et Robert [24]. Pour $n < \zeta^-$, nous avons $S_1 > 0, \ldots, S_{n-1}, S_n > 0$; nous pouvons donc écrire

$$\phi_+(u,\xi) = \sum_{n \geq 0} E\left(u^n e^{-\xi S_n} 1_{\{S_1 > 0, \ldots, S_{n-1}, S_n > 0\}}\right).$$

D'après l'indépendance des X_i, le vecteur $(X_1, X_2, \ldots, X_n)$ a même loi que $(X_n, X_{n-1}, \ldots, X_1)$; donc $(S_n, 1_{\{S_1>0,\ldots,S_{n-1},S_n>0\}})$ et $(S_n, 1_{\{S_n>S_{n-1},S_n>S_{n-2},\ldots,S_n>0\}})$ ont aussi même loi. On peut donc écrire la fonction ϕ_+ comme

$$\phi_+(u,\xi) = \sum_{n \geq 0} E\left(u^n e^{-\xi S_n} 1_{\{S_n > S_{n-1}, S_n > S_{n-2}, \ldots, S_n > 0\}}\right). \tag{17}$$

Notons ζ^+ le premier temps "record" (en considérant une échelle croissante de la marche aléatoire (S_n)); c'est aussi le temps d'atteinte du demi-plan strictement positif, puisque la valeur atteinte est strictement plus grande que la valeur 0 à l'origine. Nous avons

$$\zeta^+ = \inf\{k > 0 / S_k > 0\} = \inf\{k > 0 / S_k > S_{k-1}, \ldots, S_k > S_1, S_k > 0\}.$$

Les termes de la somme infinie de l'équation 17 valent 1 pour $n = 0$ et 0 pour $n < \zeta^+$. Nous pouvons donc écrire

$$\phi_+(u,\xi) = 1 + E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}} \sum_{n \geq \zeta^+} u^{n-\zeta^+} e^{-\xi(S_n - S_{\zeta^+})} 1_{\{S_n > S_{n-1}, S_n > S_{n-2}, \ldots, S_n > 0\}}\right). \tag{18}$$

La somme dans l'équation 18 porte sur tous les instants records de la marche aléatoire $(S_{n+\zeta^+} - S_{\zeta^+}; n \geq 0)$; d'autre part, les instants de records de la marche aléatoire $(S_{n+\zeta^+} - S_{\zeta^+})$ sont ceux de la marche aléatoire (S_n) après l'instant ζ^+. Comme par définition de ζ^+, premier temps record strictement positif, nous avons $\{S_n > S_{\zeta^+}\} \Rightarrow \{S_n > S_i, \forall i < \zeta^+\}$, nous avons

$$\begin{aligned}
&1_{\{n \geq \zeta^+, S_n > S_{n-1}, S_n > S_{n-2}, \ldots, S_n > 0\}} \\
&\quad = 1_{\{n \geq \zeta^+, S_n > S_{n-1}, S_n > S_{n-2}, \ldots, S_n > S_{\zeta^+}\}} 1_{\{n \geq \zeta^+, S_n \geq S_{\zeta^+-1}, S_n > S_{\zeta^+-2}, \ldots, S_n > 0\}} \\
&\quad = 1_{\{n \geq \zeta^+, S_n > S_{n-1}, S_n > S_{n-2}, \ldots, S_n > S_{\zeta^+}\}} \\
&\quad = 1_{\{n \geq \zeta^+, S_n - S_{\zeta^+} > S_{n-1} - S_{\zeta^+}, S_n - S_{\zeta^+} > S_{n-2} - S_{\zeta^+}, \ldots, S_n - S_{\zeta^+} > 0\}}.
\end{aligned}$$

Le lemme 1 nous dit que la loi de $S_n - S_{\zeta^+}$ est identique à celle de $S_{n-\zeta^+}$ $(n \geq \zeta^+)$. Nous pouvons donc réécrire l'espérance de la somme de l'équation 18 sous la forme

$$E\left(\sum_{n\geq\zeta^+} u^{n-\zeta^+} e^{-\xi(S_n-S_{\zeta^+})} 1_{\{n\geq\zeta^+, S_n-S_{\zeta^+}>S_{n-1}-S_{\zeta^+}, S_n-S_{\zeta^+}>S_{n-2}-S_{\zeta^+},\ldots,S_n-S_{\zeta^+}>0\}}\right)$$
$$= E\left(\sum_{n\geq\zeta^+} u^{n-\zeta^+} e^{-\xi(S_n-S_{\zeta^+})} 1_{\{S_{n-\zeta^+}>S_{n-\zeta^+-1}, S_{n-\zeta^+}>S_{n-\zeta^+-2},\ldots,S_{n-\zeta^+}>S_0\}}\right);$$

par translation de l'indice n, nous obtenons alors

$$E\left(\sum_{n\geq\zeta^+} u^{n-\zeta^+} e^{-\xi(S_n-S_{\zeta^+})} 1_{\{S_n>S_{n-1}, S_n>S_{n-2},\ldots,S_n>0\}}\right)$$
$$= \sum_{n\geq 0} E\left(u^n e^{-\xi S_n} 1_{\{S_1>0,\ldots,S_{n-1}>0,S_n>0\}}\right) = \phi_+(u,\xi).$$

Nous avons obtenu une représentation de $\phi_+(u,\xi)$ analogue à celle de $\phi_-(u,\xi)$,

$$\phi_+(u,\xi) = \frac{1}{1 - E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)}.$$

De même que $\phi_-(u,\xi)$ était bornée ainsi que son inverse sur $\Re\xi \leq 0$, $\phi_+(u,\xi)$ est bornée ainsi que son inverse sur $\Re\xi \geq 0$; de plus, comme $S_{\zeta^+} > 0$,

$$\lim_{\Re\xi\to+\infty} \phi_+(u,\xi) = 1.$$

Nous démontrons en annexe (paragraphe A.1) l'analycité de ϕ_- et ϕ_+, pour $|u| < 1$.

Nous avons montré l'existence de deux fonctions vérifiant les conditions 1. et 2. de notre théorème. Il nous reste à montrer l'unicité de la décomposition.

Soit deux autres fonctions ψ_+ et ψ_- vérifiant les propriétés (*1.*) et (*2.*). La relation 9 nous donne

$$\frac{1}{1 - uE(e^{-\xi X})} = \phi_+(u,\xi)\phi_-(u,\xi) = \psi_+(u,\xi)\psi_-(u,\xi), \quad \text{ou}$$

$$\frac{\psi_+(u,\xi)}{\phi_+(u,\xi)} = \frac{\phi_-(u,\xi)}{\psi_-(u,\xi)}, \quad |u| < 1, \Re\xi = 0. \tag{19}$$

La fonction $H(\xi)$ définie par

$$H(\xi) = \frac{\psi_+(u,\xi)}{\phi_+(u,\xi)}, \quad \Re\xi \geq 0,$$
$$H(\xi) = \frac{\phi_-(u,\xi)}{\psi_-(u,\xi)}, \quad \Re\xi \leq 0,$$

est analytique sur $\mathbb{C} - \{\Re\xi = 0\}$ et continue sur $\mathbb{C}$; elle est donc analytique sur $\mathbb{C}$ tout entier (on démontrera ces propriétés en annexe: théorème A). Cette fonction est de plus bornée, l'étant dans chacun des demi-plans $\Re(z) \leq 0$ et $\Re(z) \geq 0$; nous savons donc par le théorème de Liouville qu'elle est constante; or

$$\lim_{\Re\xi\to+\infty} H(\xi) = \lim_{\Re\xi\to+\infty} \frac{\psi_+(u,\xi)}{\phi_+(u,\xi)} = 1,$$

donc $H \equiv 1$, $\psi_+ = \phi_+$, $\psi_- = \phi_-$ et notre décomposition est bien unique.

2.3. **Loi de ζ^- (resp. ζ^+) et de S_{ζ^-} (resp. S_{ζ^+}).** Nous faisons dans ce qui suit usage de la propriété Taubérienne suivante, (voir annexe théorème 40):

Propriété 3. *Soit une suite de nombres réels $c_n \geq 0$, telle que $\sum_{n=0}^{\infty} c_n r^n$ converge pour $0 \leq r < 1$, alors*

$$\lim r \nearrow 1 \sum_{n=0}^{\infty} c_n r^n = \sum_{n=0}^{\infty} c_n,$$

où $\sum_{n=0}^{\infty} c_n$ peut être fini ou infini.

Nous allons démontrer le théorème suivant concernant le temps d'arrêt ζ^- (resp. ζ^+).

Théorème 4. *La fonction génératrice du temps d'arrêt ζ^- est donnée par*

$$E\left(u^{\zeta^-}\right) = 1 - \exp\left(-\sum_{i=1}^{\infty} \frac{u^i}{i} P\left[S_i \leq 0\right]\right), \quad \textit{et} \tag{20}$$

$$P(\zeta^- < \infty) = 1 \quad \textit{si et seulement si} \quad \sum_{i=1}^{\infty} \frac{1}{i} P\left[S_i \leq 0\right] = \infty, \tag{21}$$

auquel cas

$$E(\zeta^-) = \exp\left(\sum_{i=1}^{\infty} \frac{1}{i} P\left[S_i > 0\right]\right), \tag{22}$$

et

$$1 - E\left(e^{-\xi S_{\zeta^-}}\right) = \exp\left(-\sum_{i=1}^{\infty} \frac{1}{i} E\left(e^{-\xi S_i}; S_i \leq 0\right)\right) \tag{23}$$

Preuve. En appliquant l'unicité de la factorisation de Wiener-Hopf aux décompositions obtenues dans les équations 7 et 8, nous obtenons les deux identités

$$\phi_-(u,\xi) = \frac{1}{1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right)} = \prod_{i=1}^{\infty} \exp\left(\frac{u^i}{i} E\left(e^{-\xi S_i}; S_i \leq 0\right)\right), \quad \text{et,} \tag{24}$$

$$\phi_+(u,\xi) = E\left(\sum_{0 \leq n < \zeta^-} u^n e^{-\xi S_n}\right) = \prod_{i=1}^{\infty} \exp\left(\frac{u^i}{i} E\left(e^{-\xi S_i}; S_i > 0\right)\right). \tag{25}$$

En remarquant que $u^{\zeta^-} = 0$ pour $|u| < 1$ et $\zeta^- = \infty$, nous pouvons réécrire l'équation 24

$$1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}\right) = 1 - E\left(u^{\zeta^-} e^{-\xi S_{\zeta^-}}; \zeta^- < \infty\right) = \exp\left(-\sum_{i=1}^{\infty} \frac{u^i}{i} E\left(e^{-\xi S_i}; S_i \leq 0\right)\right); \tag{26}$$

En faisant $\xi = 0$ dans 26, comme $E\left(1; S_i \leq 0\right) = P(S_i \leq 0)$, nous obtenons l'équation 20.

Pour $|u| < 1$, en prenant une détermination convenable du logarithme, nous pouvons écrire l'équation 24 sous la forme

$$\sum_{i=1}^{\infty} \frac{u^i}{i} E\left(e^{-\xi S_i}; S_i \leq 0\right) = \log(\phi_-(u,\xi)). \tag{27}$$

L'analycité de ϕ_- et ϕ_+ entraîne que la série du premier terme de l'équation 27 converge pour $|u| < 1$. Nous pouvons donc appliquer la propriété 3 en faisant $u = 1$ dans l'équation 27, ce qui nous donne l'équation 23.

Pour la même raison, la limite de $E(u^{\zeta^-})$ s'obtient en faisant $u = 1$ dans la série du membre droit de l'équation 20. Nous notons à nouveau que $u^{\zeta^-} = 0$ pour $\zeta^- = \infty$, et comme, d'autre part, nous avons

$$\begin{aligned}\lim_{u\nearrow 1} E(u^{\zeta^-}) &= \lim_{u\nearrow 1} E(u^{\zeta^-}; \zeta^- < \infty)\\ &= \lim_{u\nearrow 1} \sum_{i=1}^{\infty} P(\zeta^- = i)u^i = \sum_{i=1}^{\infty} P(\zeta^- = i) = P(\zeta^- < \infty),\end{aligned}$$

nous obtenons que $\{P(\zeta^- < \infty) = 1\}$ si et seulement si la somme $\sum_1^\infty \frac{1}{i} P(S_i \leq 0)$ est infinie. Enfin, en faisant $\xi = 0$ dans l'équation 25, nous avons

$$E\left(\sum_{n=1}^{\zeta^- -1} u^n\right) = \exp\left(\sum_{i=1}^{\infty} \frac{u^i}{i} P\left[S_i > 0\right]\right); \tag{28}$$

$$\begin{aligned}\text{mais}\quad E\left(\sum_{n=0}^{\zeta^- -1} u^n\right) &= \sum_{k=0}^{\infty} P(\zeta^- = k) \sum_{n=0}^{k-1} u^n = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} u^n P(\zeta^- = k)\\ &= \sum_{n=0}^{\infty}\sum_{k>n}^{\infty} u^n P(\zeta^- = k) = \sum_{n=0}^{\infty} u^n P(\zeta^- > k),\end{aligned}$$

par interversion des deux sommations; nous pouvons donc réécrire l'équation 28, en utilisant la propriété 3 lorsque u tend vers 1,

$$\lim_{u\nearrow 1} \sum_{n=0}^{\infty} u^n P(\zeta^- > n) = \sum_{n=0}^{\infty} P(\zeta^- > n) = E(\zeta^-) = \exp\left(\sum_{i=1}^{\infty} \frac{1}{i} P\left(S_i > 0\right)\right). \tag{29}$$

□

3. Démonstration de la formule des valeurs extrêmes d'Iglehart-Karlin et calcul des constantes

Nous revenons à la démonstration proprement dite des formules d'Iglehart-Karlin que nous énonçons complètement:

Théorème 5. *Soit une variable aléatoire X telle que $E(X) < 0$ et $\{P(X > 0) > 0\}$ et X_n une suite de variables aléatoires indépendantes et distribuées comme X. Soit les sommes partielles $S_k = \sum_{i=1}^k X_k$ et $M(n)$ la section de valeur maximum de la marche aléatoire (S_n), telle que $M(n) = \max_{1\leq i\leq j\leq n}\{S_j - S_i\}$.*

La queue de distribution de $M(n)$ est donnée par la formule

$$\Pr\left(M(n) > \frac{\log(n)}{\lambda} + x\right) = 1 - e^{Ke^{-\lambda x}}, \tag{30}$$

où la constante λ est la solution positive de l'équation en ξ

$$E[e^{\xi X}] = 1, \tag{31}$$

et la constante K est donné par la formule

$$K = \frac{\exp\left\{-2\sum_{k=1}^{\infty}\frac{1}{k}\left(E(e^{\lambda S_k}; S_k < 0) + \Pr(S_k \geq 0)\right)\right\}}{\lambda E\left(Xe^{\lambda X}\right)}. \tag{32}$$

Preuve. Nous suivons le plan de la démonstration donné au paragraphe 1.1. Nous traitons dans le paragraphe 3.1 le cas où X est une variable continue, et dans le paragraphe 3.2 le cas où X est une variable discrète. Nous supposerons de plus que X est à support compact.

Les définitions sont celles du tableau 1.

3.1. Cas d'une variable X continue.

3.1.1. *Calcul de la queue de distribution de la somme partielle record ($R(y)$).* Considérons le temps d'arrêt ζ^+ correspondant à la première somme partielle strictement positive, de valeur $Z^+ = S_{\zeta^+}$, qui est une variable aléatoire à défaut de loi $L_{Z^+}(y) = P(0 < Z^+ \leq y)$. La transposition de l'équation 23 du cas ζ^- au cas ζ^+ nous donne

$$1 - E\left(e^{\xi Z^+}; \zeta^+ < \infty\right) = \exp\left(-\sum_{i=1}^{\infty}\frac{1}{i}E\left(e^{\xi S_i}; S_i > 0\right)\right), \tag{33}$$

que nous pouvons réécrire

$$\begin{aligned} 1 - E\left(e^{\xi Z^+}\right) &= \exp\left(-\sum_{i=1}^{\infty}\frac{1}{i}\left[E(e^{\xi S_i}) - E\left(e^{\xi S_i}; S_i \leq 0\right)\right]\right) \\ &= \exp\left(-\sum_{i=1}^{\infty}\frac{1}{i}\left(E^{\xi X}\right)^i\right)\exp\left(\sum_{i=1}^{\infty}\frac{1}{i}E\left(e^{\xi S_i}; S_i \leq 0\right)\right) \\ &= \left(1 - E(e^{\xi X})\right)\exp\left(\sum_{i=1}^{\infty}\frac{1}{i}E\left(e^{\xi S_i}; S_i \leq 0\right)\right). \end{aligned} \tag{34}$$

En choisissant λ tel que $E(e^{\lambda X}) = 1$, nous obtenons aussi $E(e^{\lambda Z^+}; \zeta^+ < \infty) = 1$. Nous montrons dans le lemme suivant qu'il existe toujours un tel λ.

Lemme 6. *Soit une variable aléatoire X telle que $E(X) < 0$ et $\{P(X > 0) > 0\}$; alors il existe un réel λ tel que $E(e^{\lambda X} = 1)$.*

Preuve. Considérons la fonction $\theta(t) = E(e^{tX})$. Nous avons $\theta(0) = 1$ et $\theta'(0) = E(X) < 0$; de plus, $\{P(X > 0) > 0\}$ implique qu'il existe ϵ et α positifs tels que $\{P(X > \epsilon) > \alpha\}$. Nous avons alors

$$E(e^{tX}) > E(e^{tX}; X > \epsilon) > e^{\epsilon t}P(X > \epsilon) = e^{\epsilon t}\alpha \to \infty \quad \text{quand} \quad t \to \infty;$$

ceci entraîne par continuité qu'il existe $t \in [0, \infty]$ tel que $\theta(t) = 1$. □

En faisant $\xi = 0$ dans l'équation 33, nous obtenons une version plus fine de 21

$$1 - P(\zeta^+ < \infty) = 1 - L_{Z^+}(\infty) = \exp\left(\sum_{i=1}^{\infty} -\frac{1}{i}P(S_i > 0)\right) \tag{35}$$

Nous considérons à présent le record absolu R de la marche aléatoire sur $[0, \infty]$ de distribution $R(y) = P(R \leq y)$. Nous avons l'équation de convolution

$$\begin{aligned} P(R \leq y) &= P(\zeta^+ = \infty) + \int_0^y P(R_{[\zeta^+,\infty]} < y - t)P(Z^+ = t)P(\zeta^+ < \infty)dt \\ &= P(\zeta^+ = \infty) + \int_0^y P(R_{[\zeta^+,\infty]} < y - t)P(Z^+ = t)dt, \end{aligned} \tag{36}$$

où $R_{[\zeta^+,\infty]}$ représente le record absolu de la marche aléatoire commençant en ζ^+; la deuxième équation s'obtient par l'identité $1_{\{Z^+>0\}} = 1_{\{\zeta^+<\infty\}}$.

Nous pouvons écrire l'équation 36 comme

$$R(y) = 1 - L_{Z^+}(\infty) + \int_0^y R(y-t)dL_{Z^+}(t), \tag{37}$$

mais dans cette équation la distribution dL_{Z^+} est une distribution à défaut ($L_{Z^+}(\infty) < 1$). Nous allons faire la substitution $R(y) = 1 - V(y)e^{-\lambda y}$; ceci nous donne

$$1 - V(y)e^{-\lambda y} = 1 - L_{Z^+}(\infty) + \int_0^y \left(1 - V(y-t)e^{-\lambda(y-t)}\right) dL_{Z^+}(t), \tag{38}$$

ou

$$V(y) = (L_{Z^+}(\infty) - L_{Z^+}(y))e^{\lambda y} + \int_0^y V(y-t)e^{\lambda t}dL_{Z^+}(t); \tag{39}$$

dans cette dernière équation, la distribution $\Lambda_{Z^+}(t) = e^{\lambda t}dL_{Z^+}(t)$ est une distribution sans défaut ($\Lambda_{Z^+}(\infty) = E(e^{\lambda Z^+}) = 1$); nous pouvons donc utiliser les théorèmes standard de la théorie du renouvellement. Le théorème du renouvellement est une conséquence du théorème de Blackwell (voir Feller [9] ou Durrett [8] pour ces deux théorèmes); ces deux théorèmes ont une formulation non dégénérée quand la distribution sur laquelle est construite le processus de renouvellement $U_i > 0, T_n = \sum_{i=1}^n U_i$ est sans défaut ($\{P(U_i) < \infty\} = 1$).

Le théorème du renouvellement s'énonce comme suit:

Théorème 7. *Soit F la fonction de répartition d'une variable aléatoire positive de moyenne μ. Supposons que h soit une fonction directement Riemann intégrable et que la fonction H soit la solution de l'équation du renouvellement*

$$H(t) = h(t) + \int_0^t H(t-x)dF(x). \tag{40}$$

Si F n'est pas arithmétique, alors

$$\lim_{t\to\infty} H(t) = \frac{1}{\mu}\int_0^\infty h(x)dx, \quad si \quad \mu < \infty.$$

Nous démontrons en annexe (paragraphe B) que, si X est absolument continue et à support compact, la fonction $(L_{Z^+}(\infty) - L_{Z^+}(y))e^{\lambda y}$ est directement Riemann intégrable (paragraphe B.1). Le théorème 7, appliqué à l'équation 39, nous donne donc:

$$\lim_{y\to\infty} V(y) = \lim_{y\to\infty} e^{\lambda y}(1 - R(y)) = \frac{\int_0^\infty e^{\lambda y}(L_{Z^+}(\infty) - L_{Z^+}(y))dy}{\int_0^\infty y d\Lambda_{Z^+}(y)} = C_1. \tag{41}$$

L'intégration par partie du numérateur de la dernière expression nous donne $\frac{1}{\lambda}(1 - L_{Z^+}(\infty))$; le dénominateur s'écrit

$$\int_0^\infty y d\Lambda_{Z^+}(y) = \int_0^\infty y e^{\lambda y} dL_{Z^+}(y) = E(Z^+ e^{\lambda Z^+}; \zeta^+ < \infty) \text{ et, donc}$$

$$C_1 = \frac{1 - L_{Z^+}(\infty)}{\lambda E(Z^+ e^{\lambda Z^+}; \zeta^+ < \infty)}, \quad \text{et} \quad R(y) \to 1 - C_1 e^{-\lambda y}. \tag{42}$$

Nous pouvons calculer le dénominateur en dérivant en ξ le premier et le dernier terme de l'équation 34, ce qui est licite compte tenu de l'analycité de la fonctions ϕ^+ dans la factorisation

de Wiener-Hopf, et en faisant $\xi = \lambda$, ce qui annule le terme contenant $(1 - E(e^{\xi X}))$. Ceci nous donne

$$E(Z^+ e^{\lambda Z^+}) = E(Xe^{\lambda X}) \exp\left(\sum_{i=1}^{\infty} \frac{1}{i} E\left(e^{\lambda S_i}; S_i \leq 0\right)\right). \tag{43}$$

3.1.2. *Calcul de la queue de distribution de la variable Q_1, maximum de la marche aléatoire avant le premier passage dans le demi-plan négatif.* Nous suivons à nouveau Iglehart [13]. Nous considérons la variable $Q = Q_1$ correspondant au maximum atteint par la marche aléatoire avant d'atteindre pour la première fois le demi-plan d'ordonnées négatives, de distribution $Q(y)$

Soit le temps d'arrêt $a_y = a_{[0,y]} = \inf\{i; S_i \notin [0, y]\}$, qui est le temps de sortie de l'intervalle $[0, y]$.

Nous remarquons que si $Q < y$, comme Q est le maximum atteint avant le premier passage dans le demi-plan négatif, la sortie de l'intervalle $[0, y]$ se fait par une valeur négative t ($t = S_{a_y} < 0$), et nous noterons

$$A^-_{(y)} = 1_{\{Q \leq y\}} S_{a_y} \quad \text{et} \quad A^-_{(y)}(t) = P(S_{a_y} \leq t, Q \leq y)$$

Nous décomposons les probabilités suivant que le record Q est supérieur ou inférieur à y. Nous écrivons donc

$$P(S_R > y) = P(S_R > y, Q \geq y) + P(S_R > y, Q < y); \tag{44}$$

d'autre part, les inégalités $Q < y \leq S_R$ indiquent que le record absolu a lieu après le temps a_y; il est équivalent de dire que le record absolu de la marche aléatoire partant de l'origine vaut y, ou que le record absolu de la marche aléatoire partant du temps a_y et du point d'abscisse négative t vaut $y - t$. En appliquant à nouveau le lemme 1, ceci s'écrit sous la forme de l'équation de convolution

$$\begin{aligned} P(S_R > y, A^-_{(y)} = t, Q < y) &= P(S_{R[a_y,\infty]} > y - t, A^-_{(y)} = t, Q < y)) \\ &= P(S_{R[a_y,\infty]} > y - t) P(A^-_{(y)} = t, Q < y); \end{aligned}$$

Nous avons d'autre part,

$$\{Q > y \Rightarrow S_R > y\}, \quad \text{et donc} \quad P(S_R > y, Q > y) = P(Q > y);$$

nous pouvons donc réécrire l'équation 44, en utilisant le lemme 1 (répliques probabilistes) pour la distribution de $S_{R[a_y,\infty]}$, sous la forme

$$P(S_R > y) = P(Q > y) + \int_{-\infty}^{0^-} P(S_R > y - t) d_t P(S_{a_y} < t, Q < y), \tag{45}$$

ou encore, sous la forme,

$$1 - R(y) = 1 - Q(y) + \int_{-\infty}^{0^-} (1 - R(y - t)) dA^-_{(y)}(t). \tag{46}$$

Quand $y \to \infty$, nous avons $dA^-_{(y)}(t) \to dZ^-(t)$, presque sûrement, avec Z^- la valeur de la marche aléatoire lors du premier record négatif (la sortie de l'intervalle des ordonnées $[0, y]$ se fait au point d'abscisse ζ^-); ceci découle du fait que $\mu = E(X) < 0$, et que, par conséquent, presque sûrement, $Q < \infty$.

En prenant la limite à l'infini, nous pouvons écrire l'équation 46 comme

$$\lim_{y\to\infty}(1 - R(y)) = \lim_{y\to\infty}(1 - Q(y)) + \lim_{y\to\infty}\left\{\int_{-\infty}^{0^-} (1 - R(y - t)) dZ^-(t)\right\};$$

en multipliant par $e^{\lambda}y$, et en utilisant le fait (équation 41) que

$$\lim_{y\to\infty}(1-R(y))e^{\lambda y} = \lim_{y\to\infty}(1-R(y-t))e^{\lambda(y-t)} = C_1,$$

nous obtenons à la limite

$$C_1 = \lim_{y\to\infty}(1-Q(y))e^{\lambda y} + \int_{-\infty}^{0^-} C_1 e^{\lambda t} dZ^-(t),$$

ce qui donne la queue de distribution de Q,

$$\lim_{y\to\infty}(1-Q(y)) = C_1(1-E(e^{\lambda Z^-}))e^{-\lambda y} = C_2 e^{-\lambda y}. \tag{47}$$

3.1.3. *Queue de distribution de $M(n)$.* Le lemme 1 nous permet d'affirmer que les marches aléatoires commençant après les instants d'échelle de records négatifs successifs q_i sont toutes des répliques de la marche aléatoire initiale; par conséquent les records $Q_1, Q_2, \ldots, Q_i, \ldots$ obtenus dans les intervalles de temps $[q_0, q_1[, [q_1, q_2[, \ldots, [q_{i-1}, q_i[, \ldots$ sont indépendants et distribués identiquement à $Q_1 = Q$, dont nous avons calculé la queue de distribution $Q(y)$ au paragraphe précédant. Considérons les m premières excursions de la marche aléatoire délimitées par les temps q_i et $M(q_m) = \max\{Q_1, Q_2, \ldots, Q_m\}$. Nous avons alors

$$P\left(M(q_m) \le \frac{\log m}{\lambda} + x\right) = \left[Q\left(\frac{\log m}{\lambda} + x\right)\right]^m;$$

en utilisant la limite obtenue dans l'équation 47, nous obtenons

$$\begin{aligned}\lim_{m\to\infty} P\left(M(q_m) \le \frac{\log m}{\lambda} + x\right) &= \lim_{m\to\infty}\left(1 - C_2 e^{-\log m} e^{-\lambda x}\right)^m \\ &= \lim_{m\to\infty}\left(1 - \frac{1}{m} C_2 e^{-\lambda x}\right)^m = e^{-C_2 e^{-\lambda x}}.\end{aligned} \tag{48}$$

Nous avons calculé la queue de distribution de $M(n)$ pour les grandes valeurs $n = q_m$; il nous reste à la calculer pour toutes les grandes valeurs de n.

Nous utilisons le théorème limite suivant de la théorie du renouvellement (voir le théorème 34 en annexe):

Théorème 8. *Soit $U_1, U_2, \ldots$ une suite de variables positives indépendantes et identiquement distribuées, et $T_i = U_1 + \cdots + U_i$ des temps de renouvellement; soit $N_t = \sup\{n; T_n \le t\}$. Alors, si $E(U_1) = \mu \le \infty$, quand $t \to \infty$, on a $N_t/t \to 1/\mu$ presque sûrement.*

Nous utiliserons ce théorème avec $U_i = q_i$, en notant $E(q_i) = E(q_1) = E(\zeta^-) = \chi$. Remarquons que $N_{q_i} = \sup\{n; T_n \le q_i\} = i$

Soit $m' = \left\lfloor \frac{n}{E(q)} \right\rfloor = \left\lfloor \frac{n}{\chi} \right\rfloor$ où $\lfloor h \rfloor$ représente la partie entière de h; le théorème 8 nous donne

$$\lim_{q_{m'}\to\infty} N_{q_{m'}}/q_{m'} \to 1/\chi;$$

En tenant compte de notre dernière remarque, qui nous donne $N_{q_{m'}} = m'$, nous obtenons $\lim_{m'\to\infty} q_{m'}/m' = \chi$, et donc,

$$\frac{q_{\left\lfloor \frac{n}{\chi} \right\rfloor}}{n} = \frac{q_{m'}}{n} \to \frac{m'\chi}{n} = \frac{\left\lfloor \frac{n}{\chi} \right\rfloor}{\frac{n}{\chi}} \to 1, \quad \text{presque sûrement, quand } n \to \infty. \tag{49}$$

Sachant que $q_1 = q = \zeta^-$, nous pouvons écrire l'équation 48 sous la forme

$$\lim_{n\to\infty} P\left(M(q_{\left\lfloor \frac{n}{\chi} \right\rfloor}) \le \frac{\log \left\lfloor \frac{n}{\chi} \right\rfloor}{\lambda} + x\right) = e^{-C_2 e^{-\lambda x}}, \tag{50}$$

et, donc, en utilisant 49,

$$\lim_{n\to\infty} P\left(M(n) \leq \frac{\log n}{\lambda} + x\right) = e^{-\frac{C_2}{\chi}e^{-\lambda x}} = e^{-\frac{C_2}{E(\zeta^-)}e^{-\lambda x}}. \tag{51}$$

En tenant compte du fait, d'une part que $Z^+ = S_{\zeta^+}$ et $Z^- = S_{\zeta^-}$, et d'autre part des équations 22, 23, 35, 42, 43, 47 et 50, nous avons donc démontré dans le cas continu la formule 1

$$\Pr\left(M(n) > \frac{\log(n)}{\lambda} + x\right) = 1 - e^{-Ke^{-\lambda x}},$$

avec λ solution positive de l'équation $E(e^{tX}) = 1$.

K est donné par

$$K = \frac{C_2}{E(\zeta^-)} = \frac{C_1(1 - E(e^{\lambda Z^-}))}{E(\zeta^-)} = \frac{(1 - L_{Z^+}(\infty))(1 - E(e^{\lambda Z^-}))}{\lambda E(Z^+ e^{\lambda Z^+}; \zeta^+ < \infty)E(\zeta^-)};$$

ce qui nous donne la formule 32,

$$K = \frac{\exp\left\{-2\sum_{k=1}^{\infty}\frac{1}{k}\left(E(e^{\lambda S_k}; S_k < 0) + \Pr(S_k \geq 0)\right)\right\}}{\lambda E\left(Xe^{\lambda X}\right)}.$$

□

3.2. **Cas d'une variable X discrète.** La démonstration de l'unicité de la factorisation de Wiener-Hopf (paragraphe 2.2), et celle de l'identité de Spitzer 23 qui en découle font usage de l'analycité de la fonction $g(\xi) = E(u^{\zeta^+}e^{-\xi S_{\zeta^+}})$ de la marche aléatoire d'incrément X. Nous utilisons comme hypothèses pour la démonstration de l'analycité de $g(\xi)$ (voir annexe A) le fait que X est à support compact. La démonstration de la décomposition de Wiener-Hopf est donc valide dans le cas où X est une variable discrète.

Pour utiliser le théorème du renouvellement (théorème 7) dans le cas continu, nous supposons (voir Annexe B) que la distribution de X est absolument continue. Nous démontrons dans ce paragraphe une version discrète du théorème du renouvellement, qui complétera la démonstration dans le cas d'un modèle discret.

Nous supposerons X arithmétique de pas δ et $|X| < A$ pour un A positif. Dans ces conditions, $Z^+ = S_{\zeta^+}$ et S_R sont des variables arithmétiques bornées positives de pas δ, et nous avons $0 < Z^+ < A$.

L'équation de convolution 36 s'écrit à présent,

$$P(R \leq y) = P(\zeta^+ = \infty) + \sum_{t=0}^{y} P\{S_{R[\zeta^+,\infty]} \leq y - t\}P(Z^+ = t), \tag{52}$$

ou, en notant $R_n = P(S_R \leq n)$, $r_n = P(S_R = n)$, $L_n = P(0 < Z^+ \leq n)$, et $l_n = P(Z^+ = n)$,

$$R_n = 1 - L_\infty + \sum_{k=0}^{n} R_{n-k}l_k. \tag{53}$$

X est supposé avoir un support discret de pas δ, et donc, nous avons $P\{X = u\} = 0$ si $u \neq z\delta$, avec $z \in \mathbb{Z}$, et de même, $r_u = l_u = 0$ si $u \neq z\delta$, avec $z \in \mathbb{Z}$.

Nous transformons l'équation 53 comme nous l'avons fait pour l'équation 37 pour obtenir une distribution sans défaut, par la transformation $R_n = 1 - V_n e^{-n\lambda}$, ce qui nous donne une équation parallèle à l'équation 39,

$$V_n = (L_\infty - L_n)e^{n\lambda} + \sum_{k=0}^{n} V_{n-k}e^{k\lambda}l_k. \tag{54}$$

Notons que, de même que dans le cas continu, nous avons

$$E(e^{\lambda Z^+}) = E(e^{\lambda X}) = 1, \quad \text{ou encore} \quad \sum_{k=0}^{\infty} e^{k\lambda} l_k = 1.$$

Nous allons à présent utiliser le théorème du renouvellement discret que nous démontrerons dans les conditions où nous l'utilisons, à savoir X à support compact, ce qui correspond dans le théorème que nous énonçons à des suites f_n et b_n n'ayant qu'un nombre fini de termes non nuls.

Théorème 9 (Théorème du renouvellement discret.). *Soit b_n et f_n deux suites à termes positifs telles que*

$$f_n \geq 0, \quad , \sum_{k=0}^{\infty} f_k = 1 \quad b_n \geq 0, \quad B = \sum_{k=0}^{\infty} b_k < \infty, \tag{55}$$
$$\textit{et telles que} \quad 0 \neq \mu = \sum n f_n < \infty.$$

Soit d'autre part l'équation de renouvellement

$$v_n = b_n + f_0 v_n + f_1 v_{n-1} + \cdots + f_n v_0. \tag{56}$$

Alors, si f_n n'est pas périodique,

$$v_n \to B\mu^{-1} \quad \textit{quand} \quad n \to \infty,$$

et, si f_n est périodique de pas δ,

$$v_{n\delta} \to B\delta\mu^{-1} \quad \textit{quand} \quad n \to \infty.$$

Preuve. Feller [9] (Volume 1) démontre ce théorème par des moyens arithmétiques et des arguments de limites de suites. Il effectue la démonstration dans le cas non périodique, puis l'étend au cas périodique.

Avec nos hypothèses de compacité sur X, on peut faire la démonstration suivante qui utilise l'intégrale de Cauchy. (On peut consulter Churchill [4] ou Flajolet et Sedgewick [10] pour cette approche).

Considérons, dans le cas non périodique, les séries génératrices associées à nos suites

$$F(z) = \sum_{n=0}^{\infty} f_n z^n, \quad B(z) = \sum_{n=0}^{\infty} b_n z^n, \quad \text{et} \quad V(z) = \sum_{n=0}^{\infty} v_n z^n \quad (z \in \mathbb{C}).$$

Comme nous l'avons mentionné précédemment, les sommes donnant $F(z)$ et $B(z)$ ne comportent qu'un nombre fini de termes; $F(z)$ et $B(z)$ sont donc des polynômes. D'autre part, nous obtenons immédiatement à partir de l'équation de renouvellement

$$V(z) = B(z) + V(z)F(z) = \frac{B(z)}{1 - F(z)}. \tag{57}$$

Nous allons montrer que le pôle $z = 1$ de $V(z)$ est le pôle de plus petit module, et qu'il est un pôle simple.

$F'(1) = \mu \neq 0$, donc $z = 1$ est un pôle simple de $V(z)$. Montrons que c'est le pôle de plus petit module.

Pour $\{|z| < 1, z \in \mathbb{C}\}$, on a $|F(z)| \leq \sum f_n |z|^n < \sum f_n = 1$.

D'autre part, pour $\{|z| = 1, z \neq 1\}$, supposons qu'il existe un terme non nul de $F(z)$ d'exposant k tel que $z^k = e^{i\theta} \neq 1$ $(0 < \theta < 2\pi)$. Nous avons

$$\Re F(z) = \sum_{i \neq k} f_i \Re z^i + f_k \cos\theta < \sum f_n = 1,$$

puisque θ est strictement compris entre 0 et 2π; donc, pour que $F(z)$ soit égal à 1, il faut que pour tous les termes non nuls de F(z) d'exposants k, on ait $z^k = 1$.

Soit alors n_0 l'exposant du plus petit terme non nul dans la série $F(z)$. On a $z^{n_0} = 1$, et donc $z = e^{2i\pi/n_0}$. La série F(z) n'étant pas périodique, il existe un terme non nul d'exposant n_1 tel que $n_1 \neq kn_0$ ($k \in \mathbb{N}$), ou encore $n_1 = n_0 q + r$, avec $r < n_0, r \in \mathbb{N}$. Nous avons alors $z^{n_1} = e^{(r/n_0)2i\pi} \neq 1$, et donc $F(z) \neq 1$.

Le polynôme $1 - F(z)$ a donc une seule racine, en $z = 1$, dans le disque $|z| \leq 1$. Les autres racines de ce polynôme sont de module strictement plus grand que 1; il existe donc $\epsilon > 0$ tel que toutes ces racines soient de module supérieur à $1 + \epsilon$.

Nous avons d'autre part $\mu = \sum n f_n = F'(1) < \infty$. En développant $B(z)$ et $1 - F(z)$ au voisinage de 1, nous obtenons

$$V(z)(1-z) = \frac{B(1) + (1-z)B'(1) + \frac{1}{2}(1-z)^2 B''(1) + \dots}{F'(1) + \frac{1}{2}(1-z)F''(1) + \dots},$$

le numérateur comme le dénominateur de la dernière expression sont des polynômes, et le résidu de $V(z) = \frac{B(z)}{1-F(z)}$ au point $z = 1$ vaut $B(1)/F'(1)$.

La fonction $h(z) = V(z) - \frac{B(1)}{(1-z)F'(1)}$ est analytique et donc bornée dans le disque $|z| \leq 1 + \epsilon$; nous pouvons maintenant appliquer la formule intégrale de Cauchy

$$\begin{aligned} v_n = [z^n]V(z) &= \int_{|z|=1+\epsilon} \frac{V(z)}{z^n} = \int_{|z|=1+\epsilon} \frac{B(1)}{z^n(1-z)F'(1)} + \int_{|z|=1+\epsilon} \frac{h(z)}{z^n} \\ &= Res_{\{z=1\}} \frac{B(1)}{z^n(1-z)F'(1)} + \mathcal{O}(1+\epsilon)^{-n} = \frac{B(1)}{F'(1)} + \mathcal{O}(1+\epsilon)^{-n} \to \frac{B(1)}{F'(1)} = B\mu^{-1}. \end{aligned}$$

Dans le cas où f_n et b_n sont des suites périodiques de période δ, nous avons

$$\begin{aligned} F(z) &= \sum f_{n\delta} z^{n\delta} = F_*(z^\delta), \quad \text{avec } F_*(z) = \sum f_{n\delta} z^n \quad \text{et} \\ B(z) &= \sum b_{n\delta} z^{n\delta} = B_*(z^\delta), \quad \text{avec } B_*(z) = \sum b_{n\delta} z^n. \end{aligned}$$

La série génératrice $V(z)$, fonction de $F(z)$ et de $B(z)$, est une fonction de z^δ et donc v_n est périodique de période δ. D'autre part, la suite $f_{*n} = f_{n\delta}$ est non périodique, et sa série génératrice est $F_*(z)$. Nous pouvons donc lui appliquer le résultat obtenu pour les suites non périodiques, et, en remarquant que $F_*(z) = F(z^{1/\delta})$, nous obtenons

$$v_{*n} = v_{n\delta} \to \frac{B_*(1)}{F_*'(1)} = \frac{B(1)\delta}{F'(1)} = B\delta\mu^{-1}.$$

□

Nous allons utiliser, dans le cas discret, le petit lemme suivant, relatif à la sommation par parties:

Lemme 10. *Soit u_i et v_i deux suites de nombres, de sommes partielles $U_n = \sum_{i=0}^n u_i$ et $V_n = \sum_{i=0}^n v_i$, et Δ l'opérateur aux différences, tel que $\Delta(G(n)) = G(n) - G(n-1)$; alors, on a:*

$$\sum_{i=1}^n U_i \Delta V_i = U_n V_n - U_0 V_0 - \sum_{i=1}^n V_{i-1} \Delta U_i. \tag{58}$$

Preuve. En effet

$$\Delta(U_i V_i) = U_i V_i - U_{i-1} V_{i-1} = \sum_{j=1}^i u_j \sum_{j=1}^i v_j - \sum_{j=1}^{i-1} u_j \sum_{j=1}^{i-1} v_j \tag{59}$$

$$= U_i v_i + v_i V_i - u_i v_i = U_i \Delta V_i + V_{i-1} \Delta U_i. \tag{60}$$

En sommant de 1 à n le premier et le dernier terme de l'équation 59, nous obtenons:

$$\sum_{i=1}^{n} \Delta(U_i V_i) = U_n V_n - U_0 V_0 = \sum_{i=1}^{n} U_i \Delta V_i + \sum_{i=1}^{n} V_{i-1} \Delta U_i,$$

ce qui nous donne l'équation 58. □

Dans le cas discret, nous pouvons alors énoncer:

Théorème 11. *Si X est une variable discrète de pas δ, la queue de distribution du record absolu S_R de la marche aléatoire (S_i) est donnée par*

$$R(y) \to 1 - C_1 e^{-\lambda y},$$

$$\text{avec} \quad C_1 = \frac{(1 - L_{Z^+}(\infty))\delta}{(e^{\lambda\delta} - 1) E(Z^+ e^{\lambda Z^+}; \zeta^+ < \infty)}. \tag{61}$$

Nous allons appliquer le théorème 9 à l'équation 54

$$V_n = (L_\infty - L_n) e^{n\lambda} + \sum_{k=0}^{n} V_{n-k} e^{k\lambda} l_k, \tag{62}$$

$$\text{avec} \quad f_k = e^{k\lambda} l_k, \quad v_k = V_k, \quad \text{et} \quad b_n = (L_\infty - L_n) e^{n\lambda}.$$

Remarquons que, X étant à support discret de pas δ, nous avons

$$\{f_i = 0, \quad \forall i \neq n\delta\},$$

$$\{R_{n\delta+i} = P(S_R \leq n\delta + i) = R_{n\delta}, \quad \forall i < \delta\}.$$

Nous pouvons donc nous limiter à ne considérer que les termes de rang $i\delta$ dans l'équation 62, ce qui nous donne une équation de la forme

$$V_{n\delta} = (L_\infty - L_{n\delta}) e^{n\delta\lambda} + \sum_{k=0}^{n} V_{(n-k)\delta} e^{k\delta\lambda} l_{k\delta}, \tag{63}$$

Le théorème du renouvellement discret nous donne:

$$V_{n\delta} \to \sum_{i=0}^{\infty} (L_\infty - L_{n\delta}) e^{n\delta\lambda} \delta \mu^{-1}. \tag{64}$$

Nous sommons par partie cette dernière somme, en utilisant le lemme 10, avec

$$u_0 = L_\infty, \quad u_i = -l_{n\delta} \ (i \geq 1), \quad v_i = e^{i\delta\lambda}, \quad \text{et} \quad V_i = \frac{e^{(i+1)\delta\lambda} - 1}{e^{\delta\lambda} - 1}$$

En remarquant d'une part que $L_{n\delta} = L_\infty$ pour $n\delta > M$, avec M borne supérieure du support compact de X, et d'autre part que $\sum_{i=1}^{\infty} l_{i\delta} = L_\infty$, que $\sum_{i=1}^{\infty} l_{i\delta} e^{i\lambda\delta} = 1$, et que $L_0 = l_0 = P(Z^+ = 0) = 0$, nous avons:

$$\begin{aligned}
\sum_{i=0}^{\infty} (L_\infty - L_{n\delta}) e^{n\delta\lambda} &= (L_\infty - L_0) + \sum_{i=0}^{\infty} (L_\infty - L_{n\delta}) e^{n\delta\lambda} \\
&= L_\infty - \left[(L_\infty - L_{n\delta}) \frac{e^{(n+1)\lambda\delta} - 1}{e^{\lambda\delta} - 1} \right]_0^{\infty} + \sum_{i=0}^{\infty} \frac{e^{i\lambda\delta} - 1}{e^{\lambda\delta} - 1} l_{i\delta} \\
&= L_\infty - L_\infty + \frac{1 - L_\infty}{e^{\lambda\delta} - 1}.
\end{aligned}$$

Nous obtenons ainsi 61. Nous remarquons qu'en faisant tendre δ vers 0 dans l'équation 61, nous retrouvons l'équation 42

Remarque. Les calculs donnant les queues de distribution des variables Q et $M(n)$ se déduisent immédiatement, dans le cas discret, en remplaçant les intégrales par des sommes, de ceux effectués dans le cas continu; ils nous redonnent les formules 47 et 51.

Note 12 (Une démonstration directe du théorème 11.).

Nous avons démontré le théorème 11 en suivant l'approche d'Iglehart et de Karlin, c'est à dire en faisant le changement de variable $V_n = 1 - R_n e^{-\lambda n}$, puis en considérant la solution asymptotique d'une équation de renouvellement sur une distribution sans défaut. On peut faire une démonstration plus directe qui est peut-être plus intuitive.

On part à nouveau de l'équation 53,

$$R_n = 1 - L_\infty + \sum_{k=0}^{n} R_{n-k} l_k.$$

En passant aux séries génératrices, on obtient, dans le cas non périodique

$$\sum R_n z^n = (1 - L_\infty) \sum z^n + \sum l_k z^k \sum R_n z^n,$$

ou, encore

$$R(z) = (1 - L_\infty)\frac{1}{1-z} + l(z)R(z) = \frac{1 - L_\infty}{(1-z)(1-l(z))}. \tag{65}$$

On remarque alors que

$$l(e^\lambda) = \sum_{i=0}^{\infty} l_i e^{\lambda i} = E(e^{\lambda Z^+}) = 1,$$

et par un raisonnement analogue au raisonnement fait pour la démonstration du théorème du renouvellement discret, on montre que e^λ est, pour $R(z)$, en dehors du pôle $z = 1$, le pôle de plus petit module; c'est un pôle simple puisque $l'(e^\lambda) = E(Z^+ e^{\lambda Z^+}) \neq 0$. Le résidu de $R(z)$ en 1 vaut $(1 - L_\infty)/(1 - l(1)) = 1$. Le résidu en e^λ vaut $\frac{1-L_\infty}{(1-e^\lambda)l'(e^\lambda)}$, et on obtient donc directement par le théorème de Cauchy

$$R_n = 1 - \frac{1 - L_\infty}{(e^\lambda - 1)E(Z^+ e^{\lambda Z^+})} e^{-\lambda n} + o(e^{-\lambda n}).$$

Dans le cas discret de pas δ, on fait apparaître le pôle $e^{\lambda\delta}$ et le facteur δ apparaît dans le calcul de $l'(1)$.

Remarque. Les calculs donnant les queues de distribution des variables Q et $M(n)$ se déduisent immédiatement, dans le cas discret, en remplaçant les intégrales par des sommes, de ceux effectués dans le cas continu; ils nous redonnent les formules 47 et 51.

3.3. **Deux autres résultats asymptotiques.** Karlin et Dembo donnent deux autres résultats asymptotiques intéressants.

- Un résultat sur la longueur des segments de score maximum.
- Un résultat utilisant une variable U_i couplée à la variable X_i qu'on peut appliquer pour trouver la distribution limite des lettres dans le segment de score maximum.

Ces résultats sont basés sur des propriétés des martingales de Wald simple ou double et sur le théorème d'échantillonnage optionnel.

Nous définissons les martingales.

Définition 13 (Martingales.). *Un processus stochastique* $\{X_n; n = 0, 1, \ldots\}$ *est une martingale si* $\forall n$,

1. $E(|X_n|) < \infty$, *et*
2. $E(X_{n+1}|X_0, \ldots, X_n) = X_n$.

Avec $\phi(t) = E(e^{tX})$ et $S_m = \sum_1^m X_i$, la martingale simple de Wald s'écrit

$$P_m = \frac{e^{tS_m}}{[\phi(t)]^m}. \tag{66}$$

En notant $W_k = \sum_{i=1}^k U_i$, et $\psi(t, \xi) = E(e^{tX+\xi U})$, la martingale de Wald double s'écrit

$$H_m = \frac{e^{tS_m+\xi W_m}}{[\psi(t,\xi)]^m}. \tag{67}$$

On montre facilement que P_m est une martingale; en effet, X_{m+1} est indépendant de $\{X_1, \ldots, X_m\}$, et donc

$$\begin{aligned} E(P_{m+1}) &= \frac{E(e^{t(X_1+\cdots+X_m+X_{m+1})}|X_1, \ldots X_m)}{E(e^{tX})^{m+1}} \\ &= \frac{E(e^{t(X_1+\cdots+X_m)}|X_1, \ldots X_m)E(e^{tX_{m+1}})}{E(e^{tX})^{m+1}} \\ &= \frac{e^{t(X_1+\cdots+X_m)}}{E(e^{tX})^m} = P_m. \end{aligned}$$

nous allons utiliser la variante suivante du théorème d'échantillonnage optionnel qui s'applique aux martingales à variations bornées.

Théorème 14 (Théorème d'échantillonnage optionnel.). *Soit P_n une martingale définie sur une suite X_n (les X_i sont i.i.d, et P_n ne dépend que de $X_1, \ldots, X_n$); si T est un temps d'arrêt relatif aux X_n, et si*

1. $E(T) < \infty$,
2. *et il existe une constante A telle que*

$$E(|P_{n+1} - P_n| \text{ si } X_0, \ldots, X_n) \leq A \qquad \text{pour } n < T,$$

alors, $E(P_T) = E(X_0)$.

Preuve. On peut consulter Karlin et Taylor [15] p. 260 pour la démonstration. □

Nous allons considérer le temps d'arrêt L correspondant à la première sortie de l' intervalle $[0, y]$, et nous ferons tendre y vers l'infini. Nous avons

$$L(y) = \{\min k,\ S_k > y \ \text{ ou } \ S_k < 0\}.$$

Nous définissons également un indicateur $I(y)$ disant si l'on sort de l'intervalle $[0, y]$ par le haut ($S_L > y$) ou par le bas ($S_L < 0$),

$$I(y) = 1 \ \text{ si } \ S_L > y;$$
$$I(y) = 0 \ \text{ si } \ S_L < 0;$$

Nous allons considérer le temps d'atteinte ρ du record absolu S_R, et utiliser l'inégalité de Chebyshev.

Cette inégalité s'écrit, pour une fonction $f(x)$ positive et strictement croissante sur $[0, \infty]$,

$$\Pr(|X| \geq u) \leq \frac{E[f(X)]}{f(u)}.$$

Nous allons appliquer cette inégalité à la fonction $f(x) = e^{tx}, t \geq 0$, qui vérifie les conditions posées, et qui vérifie de plus $f(0) = 1$.

Nous avons alors, pour ρ,

$$\Pr(\rho = k) \leq \Pr(S_k > 0) \leq E(e^{tS_k}) = E(e^{tX})^k.$$

La fonction $E(e^{tX})$ a pour dérivée seconde $E(X^2 e^{tX}) > 0$; elle est donc convexe, et vaut 1 pour $t = 0$ et pour $t = \lambda$. Nous avons donc $E(e^{tX}) < 1$ pour $0 < t < \lambda$, et donc

$$\Pr(\rho = k) < e^{-bk}$$

pour un $b > 0$. Nous en déduisons que la queue de probabilité de ρ décroît exponentiellement vite.

Revenons à notre variable $L(y)$; nous avons

$$\rho > y \Rightarrow L > y,$$

ce qui s'écrit sous forme de probabilités

$$\Pr(L > y) \leq \Pr(\rho > y) = \mathcal{O}(e^{-y}).$$

Nous en déduisons immédiatement

$$E(L) = \int \Pr(L > x) dx < \infty.$$

En prenant comme temps d'arrêt le temps L, nous avons donc vérifié la première condition du théorème d'arrêt optionnel 14. La deuxième condition se déduit immédiatement du fait que nos variables X sont à support bornés. En appliquant ce théorème , nous obtenons donc

$$E(e^{tS_L - L\log(E[e^{tX}])}) \equiv E(e^{tX - \log(E[e^{tX}])}) = 1. \tag{68}$$

Nous énonçons alors le premier résultat asymptotique annoncé dans ce paragraphe.

Théorème 15. *Soit $I_k(y)$ l'indicateur $I(y)$ défini ci-dessus et relatif à la section de marche aléatoire partant du kième temps d'échelle décroissante q_k. Si $T_k(y)$ est le premier temps dans la succession des sections tel que $I_k(y) = 1$, alors*

$$\frac{L_k(y)}{y} \to \frac{1}{\omega} = \frac{1}{E(Xe^{\lambda X})},$$

où λ est la seule racine positive de l'équation $E(e^{tX}) = 1$.

Preuve. (Esquisse). On se référera à Karlin et Dembo [20] (lemmes 2 et 3) pour une démonstration détaillée.

Le lemme 2 de [20] donne $\Pr(I(y) = 1)e^{\lambda y} \geq \delta > 0$.

Dans le lemme 3 de [20], Karlin et Dembo utilisent les moments de la martingale en dérivant quatre fois l'équation 68, et en faisant $t = \lambda$. La première dérivation donne

$$E[(S_L - \omega L)e^{\lambda S_L}] = 0, \quad \omega = E(Xe^{\lambda X}).$$

On obtient, en décomposant suivant les valeurs de l'indicateur $I(y)$,

$$E(Le^{\lambda S_L}) = \mathcal{O}(y);$$

puis, on obtient une inégalité quadratique, qui donne

$$E(L^2 e^{\lambda S_L}) = \mathcal{O}(y^2).$$

On obtient ensuite

$$E\left[(S_L - \omega L)^4 e^{\lambda S_L}\right] = \mathcal{O}(y^2).$$

On en déduit, après utilisation du lemme 2, et en tenant compte du fait que $S_L - y$ est borné, conditionnellement à $I(y) = 1$,

$$E\left[\left(\frac{L(y)}{y} - \frac{1}{\omega}\right)^4 \mid I(y) = 1\right] = O\left(\frac{1}{y^2}\right) \qquad (y \to \infty).$$

□

En considérant à présent la martingale double H_m définie en 67, Karlin et Dembo obtiennent le second théorème annoncé.

Théorème 16. *Soit U_i une suite de variables aléatoires bornées et i.i.d telles que U_i peut dépendre de X_i, mais est indépendant de X_j, $j \neq i$. Soit $W_j = \sum_{i=1}^{j} U_i$ et*

$$W_k(y) = W_{T_k(y)} - W_{q_{k-1}},$$

où $T_k(y)$ est le temps de la première sortie de l'intervalle $[0, y]$ par le haut, obtenu lors de la kième excursion relative au temps d'échelle décroissante. Alors, en excluant un ensemble de mesure nulle,

$$\frac{W_k(y)}{L_k(y)} \to v \quad p.s. \text{ quand } y \to \infty, \tag{69}$$

où $v = E(Ue^{\lambda X})$.

Preuve. La démonstration suit le déroulement de la démonstration faite pour le théorème 15, en dérivant l'identité

$$E(e^{tS_L + \xi W_L - L\log(E[e^{tX+\xi W}])}) \equiv 1,$$

et en faisant après chaque étape de dérivation $t = \lambda$ et $\xi = 0$. □

On peut appliquer ce résultat au cas où l'on considère un ensemble de lettres $\{a_i\}$, avec un score s_i et une probabilité p_i associés.

On considérera la lettre a_i, et on fera $U_j = 1$ si la lettre à la position j est a_i et $U_j = 0$ dans le cas contraire. La variable W_k du théorème 16 compte alors le nombre d'occurrence de la lettre a_i dans le segment $[q_k, T_k(y)]$; asymptotiquement, le théorème 16 donne la fréquence d'apparition de la lettre a_i dans le segment de score maximum, qu'on peut exprimer par le corollaire suivant.

Corollaire 17. *Soit un alphabet $a_1, \ldots, a_r$, et un score s_i et une probabilité p_i associés à chaque lettre a_i. Asymptotiquement, la fréquence d'apparition f_i de la lettre a_i dans le segment de score maximal vaut $f_i = p_i e^{\lambda s_i}$.*

Ce corollaire s'applique immédiatement au cas de l'alignement de deux séquences, où les lettres r_i d'une requête et c_j d'une cible ont les probabilités d'apparition respectives q_i et p_j, et où le score d'appariement de deux lettres r_i et c_j vaut s_{ij} Nous obtenons alors la distribution limite $p_j^\star$ dans le segment de la cible correspondant à l'alignement optimal comme

$$p_j^\star = \sum_i q_i p_j e^{\lambda s_{ij}},$$

où λ est donné par $\sum_{i,j} q_i p_j e^{\lambda s_{ij}} = 1$.

En particulier, dans le cas où la séquence requête est composée d'une seule lettre i_0, $(q_{i_0} = 1;\ q_i = 0, i \neq i_0)$, nous obtenons

$$p_j^\star = p_j e^{\lambda s_{i_0 j}}. \tag{70}$$

3.4. **Rapidité de la convergence de la série intervenant dans le calcul de la constante K.** La constante K de Karlin est donnée par la formule 32

$$K = \frac{\exp\left\{-2\sum_{k=1}^{\infty} \frac{1}{k}\left(E(e^{\lambda S_k}; S_k < 0) + \Pr(S_k \geq 0)\right)\right\}}{\lambda E\left(Xe^{\lambda X}\right)}.$$

Nous allons nous intéresser à la rapidité de convergence de la série apparaissant comme exposant au numérateur de cette expression, qui peut s'écrire

$$S = \sum_{k=1}^{\infty} \frac{1}{k} \Big(E(e^{\lambda S_k}; S_k < 0) + \Pr(S_k \geq 0) \Big) = \sum_{k=1}^{\infty} \frac{1}{k} E(e^{\lambda S_k^-}), \qquad \text{avec } S_k^- = \min(S_k, 0). \tag{71}$$

Nous allons montrer que cette série converge géométriquement vite (ce résultat de vitesse de convergence est annoncé sans démonstration par Karlin et Dembo [14]).

Nous utiliserons le théorème sur les grandes déviations suivant (voir Durrett p.70-73)

Théorème 18 (Théorème des grandes déviations.). *Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes et identiquement distribuées, telle qu'il existe un $\theta > 0$ pour lequel*

$$\phi(\theta) = E(e^{\theta X_i}) < \infty. \tag{72}$$

Alors, si $\kappa(t) = \log \phi(t)$, $S_n = \sum_{k=1}^{n} X_k$, $\mu = E(X_i)$, et si a est une constante strictement supérieure à μ, $a > \mu$, on a, pour n assez grand,

$$P(S_n \geq na) \leq e^{-n\{at - \kappa(t)\}}, \tag{73}$$

et on a, pour $t > 0$ petit, $at - \kappa(t) > 0$.

On peut réécrire l'équation 73 de ce théorème en disant qu'il existe $\epsilon > 0$ tel que

$$P(S_n \geq na) \leq e^{-n\epsilon}. \tag{74}$$

Nous allons alors démontrer:

Théorème 19. *La série permettant de calculer la constante K de Karlin converge géométriquement vite.*

Preuve. Considérons le terme $E(e^{\lambda S_k^-})$ de l'équation 71; nous pouvons l'écrire sous la forme

$$E(e^{\lambda S_k^-}) = \int_{-\infty}^{\alpha} e^{\lambda S_k^-} d\mathcal{P} + \int_{\alpha}^{0} e^{\lambda S_k^-} d\mathcal{P} + \int_{0}^{-\infty} e^{\lambda S_k^-} d\mathcal{P}, \tag{75}$$

où $\mathcal{P}$ est la probabilité induite par la suite X_i pour la variable aléatoire S_k, et où $\alpha = -\log(k)/\lambda < 0$ est solution en x de l'équation $e^{\lambda x} = 1/k$. Remarquons que

- pour $S_k^- \in]-\infty, \alpha]$, on a $e^{\lambda S_k^-} < e^{\lambda \alpha} = \frac{1}{k}$;
- pour $S_k^- \in [\alpha, 0]$, on a $e^{\lambda S_k^-} < 1$;
- et pour $S_k^- \geq 0$, on a $S_k^- = 0$, et donc $e^{\lambda S_k^-} = 1$;

ceci nous donne

$$E(e^{\lambda S_k^-}) < \int_{-\infty}^{\alpha} \frac{1}{k} d\mathcal{P} + \int_{\alpha}^{\infty} 1 d\mathcal{P} < \frac{1}{k} + P(S_k^- \geq \alpha) = \frac{1}{k} + P(S_k \geq \alpha),$$

puisque l'on a $\alpha < 0$. C'est pour borner la probabilité $P(S_k \geq \alpha)$ que nous allons utiliser le théorème des grandes déviations.

Remarquons que le théorème des grandes déviations ne fait pas d'hypothèses sur le signe de l'espérance μ de X_i; l'espérance de X_i étant négative pour toute nos démonstrations concernant les résultats de Karlin, nous avons $\mu < 0$, et la fonction $g(x) = \frac{\mu}{2}x + \frac{\log x}{\lambda}$ décroît jusqu'à $-\infty$ quand x croît de $-2/\lambda\mu$ à $+\infty$; il existe donc $x_0 > 0$ tel que $\forall x > x_0$ on ait $g(x) < 0$, et, de manière équivalente, tel que, $\forall k > x_0$, on ait $k\frac{\mu}{2} < -\frac{\log k}{\lambda}$.

Nous avons donc, pour $k > x_0$,

$$P(S_k \geq \alpha) = P\left(S_k \geq -\frac{\log k}{\lambda}\right) < P\left(S_k > k\frac{\mu}{2}\right) < e^{-k\epsilon},$$

en appliquant le théorème des grandes déviations (théorème 18) avec $\mu < a = \mu/2 < 0$, et sous la forme de l'équation 74. Nous avons obtenu

$$E(e^{\lambda S_k^-}) < \frac{1}{k} + e^{-k\epsilon} < \frac{2}{k}, \tag{76}$$

la dernière inégalité étant vérifiée pour k assez grand. La convergence de la série est donc au moins géométrique.

□

3.5. **Probabilité d'obtenir plusieurs alignements de scores supérieurs à un score donné.** Nous évaluons dans ce paragraphe la probabilité d'obtenir i alignements de score supérieur à un score y. Nous suivons pour ceci Karlin et Altschul [18] pour obtenir une formule qui est une généralisation de la formule 30 . Avec λ et K donnés par les formules 31 et 32, nous normalisons les scores en écrivant

$$S' = \lambda S - \log(Kn).$$

La formule 30 s'écrit alors

$$P(S' \geq x) \approx 1 - e^{-e^{-x}}. \tag{77}$$

Nous avons alors.

Lemme 20. *La probabilité d'obtenir r sections de la marche aléatoire d'incrément X de scores normalisés respectifs $S'_1, S'_2, \ldots, S'_r$, tels que $S'_1 \geq S'_2 \geq \cdots \geq S'_r$ est bien approchée par l'équation*

$$\Pr(S'_r \geq x) \approx 1 - e^{-e^{-x}} \sum_{k=0}^{r-1} \frac{e^{-kx}}{k!}. \tag{78}$$

Preuve. La densité de la variable aléatoire du score S' est approchée pour S' assez grand par $f(x) = \exp(e^{-x})e^{-x} = \exp(e^{-x} - x)$. La densité conjointe des $S'_1, S'_2, \ldots, S'_r$ est alors approchée par

$$f(x_1, \ldots, x_r) = \exp\left(-e^{-x_r} - \sum_{k=1}^{r} x_k\right),$$

où on a $x_1 \geq x_2 \geq \ldots x_r$. On peut alors évaluer par intégration la queue de distribution de S'_r. En effectuant une intégration par partie, nous écrivons

$$\Pr(S'_r \geq x) = \int_{x_1 \geq x_2 \geq \cdots \geq x_r = x} f(x_1, \ldots x_r) dx_1 dx_2 \ldots dx_r \tag{79}$$

$$= \int_x^\infty dx_r \int_{x_r}^\infty dx_{r-1} \cdots \int_{x_2}^\infty \exp\left(-e^{-x_r} - \sum_{k=1}^{r} x_k\right) dx_1 \tag{80}$$

$$= \int_x^\infty \frac{e^{-e^{-t}} e^{-rt}}{(r-1)!} dt \tag{81}$$

$$= \left[e^{-e^{-t}} e^{-(r-1)t}\right]_x^\infty - \int_x^\infty \frac{e^{-e^{-t}} e^{-(r-1)t}}{(r-2)!} \tag{82}$$

$$= 1 - e^{-e^{-x}} \sum_{k=0}^{r-1} \frac{e^{-kx}}{k!}. \tag{83}$$

Ceci nous donne le résultat annoncé. □

Annexe A. Analycité de ϕ_+, ϕ_- et H dans la décomposition de Wiener-Hopf

Nous allons démontrer l'analycité des fonctions ϕ^+ et ϕ^-, puis celle de H, fonctions utilisées dans la décomposition de Wiener-Hopf.

A.1. **Continuité et analycité de ϕ_+ et de ϕ_-.** Nous allons démontrer l'analycité de ϕ_+; la démonstration de l'analycité de $\phi-$ est identique. Nous pouvons écrire

$$\phi_+(u,\xi) = \frac{1}{1-g(u,\xi)},$$

avec $g(\xi) = E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)$; la continuité et l'analycité de $g(\xi)$ entraîneront celles de ϕ_+. Nous démontrerons la continuité et l'analycité de $g(\xi)$ en deux lemmes séparés; nous ferons usage pour la démonstration de l'analycité de $g(\xi)$ d'un théorème de Khoan [21] relatif à la dérivation sous le signe intégral que nous énonçons.

Théorème 21. *Soit μ une mesure sur un espace topologique séparé et localement compact Ω, f une fonction μ intégrable, et $F = \int_\Omega f(x,\xi)d\mu(x)$. Supposons que Ξ soit un ouvert de $\mathbb{R}$ ou de $\mathbb{C}$ et que les conditions suivantes soient réalisées:*

1. *pour μ-presque tout x, la fonction $\xi \to f(x,\xi)$ est continûment dérivable sur Ξ;*
2. *pour tout $\xi \in \Xi$, la fonction $x \to \frac{\partial f}{\partial \xi}(x,\xi)$ est μ-mesurable;*
3. *il existe une fonction μ-intégrable, h, telle que*

$$\left|\frac{\partial f}{\partial \xi}(x,\xi)\right| \leq h(x)$$

pour μ-presque tout $x \in \Omega$.

Alors, F est continûment dérivable sur Ξ et

$$\frac{\partial F}{\partial \xi}(\xi) = \int_\Omega \frac{\partial f}{\partial \xi}(x,\xi)d\mu(x).$$

Preuve. La démonstration utilise le théorème des accroissements finis (voir Gourdon [11]) et le théorème de la convergence dominée de Lebesgue (voir Deheuvels [6] ou Durrett [8] et Dieudonné, Élements d'Analyse [7]); On peut aussi consulter Whittaker et Watson [29] (p.92) qui ne fait pas intervenir la théorie de la mesure, mais il faut que f soit une fonction continue de x. □

Remarque. Dans les cas que nous considérons, nous avons $\Omega = \mathbb{R}$, ou $\Omega = \mathbb{Z}$, qui sont séparés et localement compacts.

Nous avons donc à démontrer le lemme suivant

Lemme 22. *Dans les conditions énoncées au théorème 2 la fonction $g(\xi) = E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)$ est continue.*

Preuve. Nous avons

$$\begin{aligned} |g(\xi+d\xi) - g(\xi)| &= \left|E\left(u^{\zeta^+} e^{-(\xi+d\xi)S_{\zeta^+}}\right) - E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)\right| \\ &= \left|E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\left(e^{-d\xi S_{\zeta^+}} - 1\right)\right)\right| \leq E\left|u^{\zeta^+} e^{-\xi S_{\zeta^+}}\left(e^{-d\xi S_{\zeta^+}} - 1\right)\right|. \end{aligned} \tag{84}$$

Nous pouvons appliquer l'inégalité de Cauchy-Schwartz $E|XY| \leq \sqrt{E|X|^2}\sqrt{E(|Y|^2}$ à l'équation 84, ce qui nous donne

$$|g(\xi+d\xi) - g(\xi)| \leq \sqrt{E\left|u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right|^2}\sqrt{E\left|e^{-d\xi S_{\zeta^+}} - 1\right|^2} \leq \sqrt{E\left(\left|e^{-d\xi S_{\zeta^+}} - 1\right|^2\right)},$$

puisque $|u| \leq 1$, $\Re\xi \geq 0$ et $S_{\zeta^+} \geq 0$. Pour z complexe, nous avons

$$|e^z - 1| = \left|\sum_{n=1}^{\infty} \frac{z^n}{n!}\right| \leq \sum_{n=1}^{\infty} \frac{|z|^n}{n!} = e^{|z|} - 1,$$

et d'autre part la fonction de x réel $t(x) = 2x - e^x + 1$ est croissante sur $[0, \log(2)]$, et vaut 0 pour $x = 0$; donc, pour $|z| < \log 2$, nous avons $|e^z - 1| \leq e^{|z|} - 1 < 2|z|$; comme nous avons supposé X à support compact, il existe une constante $A > 1$ telle que $|S_{\zeta^+}| < \max|X| < A$; par conséquent, pour $|d\xi| < \frac{\log(2)}{A}$, nous pouvons majorer $|g(\xi + d\xi) - g(\xi)|$ par

$$|g(\xi + d\xi) - g(\xi)| \leq \sqrt{E\left(2|d\xi S_{\zeta^+}|\right)^2} \leq 2A|d\xi|;$$

La continuité de $g(\xi)$ est démontrée. □

Lemme 23. *Dans les conditions énoncées au théorème 2 la fonction* $g(\xi) = E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)$ *est analytique.*

Preuve. Nous voulons démontrer que

$$\phi_+(u, \xi) = \frac{1}{1 - E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)}.$$

est analytique.

Une condition nécessaire et suffisante d'analycité d'une fonction est qu'elle soit continûment dérivable (voir Churchill [4]). C'est ce que nous allons démontrer pour la fonction $g(\xi) = E\left(u^{\zeta^+} e^{-\xi S_{\zeta^+}}\right)$ en montrant qu'elle vérifie les conditions du théorème A , ce qui entraînera l'analycité de la fonction ψ_+. Nous suivons en partie Feller [9] pour cette démonstration.

Considérons la distribution conjointe de ζ^+ et de S_{ζ^+}

$$P\{\zeta^+ = n, S_{\zeta^+} \in I\} = H_n\{I\},$$

pour tout I Borélien de $\mathbb{R}^+$; nous avons

$$H_n\{I\} = P\{S_1 \leq 0, \ldots, S_{n-1} \leq 0, S_n \in I\}, \tag{85}$$

avec $H_0\{I\} = 0$. Nous pouvons écrire $g(\xi)$ sous la forme

$$g(\xi) = \sum_{n=1}^{\infty} u^n \int_0^{\infty} e^{-\xi x} H_n(dx) = \sum_{n=1}^{\infty} \int_0^{\infty} e^{-\xi x} u^n H_n(dx); \tag{86}$$

l'intégrande étant positive, nous pouvons intervertir les signes somme et intégrale, ce qui nous donne

$$g(\xi) = \int_0^{\infty} \sum_{n=1}^{\infty} e^{-\xi x} u^n H_n(dx) = \int_0^{\infty} e^{-\xi x} \sum_{n=1}^{\infty} u^n H_n(dx);$$

sur $\mathbb{R}^+$ la mesure $\mu(dx) = \sum_{n=1}^{\infty} u^n H_n(dx)$ vérifie

$$\mu[0, +\infty] = \sum u^n P(\zeta^+ = n) < \frac{P(\zeta^+ < \infty)}{1 - u},$$

et donc $\mu[0, +\infty] < \infty$ pour $u < 1$; nous devons à présent vérifier les conditions du théorème 21 avec $f(x, \xi) = e^{-x\xi}$.

- Condition 1: elle est immédiatement vérifiée.

- Condition 2: considérons l'intégrale

$$T(\xi) = \int_0^\infty \frac{\partial f}{\partial \xi}(x,\xi)\mu(dx) = \int_0^\infty xe^{-\xi x} \sum_{n=1}^\infty u^n H_n(dx);$$

nous pouvons majorer $T(\xi)$ par

$$T(\xi) \leq \int_0^\infty xe^{-\xi x} \frac{1}{1-u} dx = \frac{1}{\xi^2}\frac{1}{1-u}.$$

L'intégrande est positive et $T(\xi)$ est bornée; la fonction $x \to \frac{\partial f}{\partial \xi}(x,\xi)$ est donc μ-mesurable;

- Condition 3: elle est immédiatement vérifiée, puisque $\frac{\partial f}{\partial \xi}(x,\xi)$ est une fonction à valeurs partout positives ou nulles, et qu'on peut donc utiliser le résultat obtenu pour la démonstration de la condition 2.

Toutes les conditions d'application du théorème 21 sont vérifiées; la fonction $g(\xi)$ est analytique pour $u < 1$, et il s'en suit que ϕ_+ l'est aussi, puisque $|g(u,\xi)| < 1$. □

A.2. **Continuité et analycité de H.** L'équation 19 s'écrit

$$H(\xi) = \frac{\psi_+(u,\xi)}{\phi_+(u,\xi)} = \frac{\phi_-(u,\xi)}{\psi_-(u,\xi)}, \quad |u| < 1, \Re\xi = 0.$$

La continuité de H découle immédiatement de celles de ϕ_+ et ψ_+ d'une part, et de ϕ_- et ψ_- d'autre part, dans leurs demi-plans respectifs ($\{\Re\xi \leq 0\}, \{\Re\xi \geq 0\}$) et de cette dernière équation pour $\{\Re\xi = 0\}$.

Pour démontrer l'analycité de H, nous allons utiliser le théorème de Morera (voir Bak et Newman [2] ou Churchill [4]), que nous énonçons.

Théorème 24 (Morera). *Si une fonction F est continue dans un domaine simplement connexe D et si, pour tout contour fermé C intérieur à D, on a*

$$\int_C f(z)dz = 0, \tag{87}$$

alors f est analytique sur D tout entier.

Nous avons alors:

Lemme 25. *La fonction H est analytique sur $\mathbb{C}$ tout entier.*

Preuve. Nous savons que H est continue sur $\mathbb{C}$ tout entier, et qu'elle est analytique d'une part sur $\{\Re\xi \leq 0\}$, puisque les fonctions ϕ_- le sont dans ce domaine, et d'autre part sur $\{\Re\xi \geq 0\}$, puisque les fonctions ϕ_- le sont dans cet autre domaine. Soit alors un contour quelconque C du plan complexe; nous avons les deux cas suivants:

1. C ne coupe pas l'axe imaginaire, et dans ce cas, C est entièrement contenu dans l'un des deux demi-plans $\{\Re\xi \leq 0\}$ et $\{\Re\xi \geq 0\}$; dans ce cas, H étant analytique dans chacun de ces demi-plans, l'intégrale $\int_C H(z)dz$ vaut 0, par application du théorème de Cauchy.
2. C coupe l'axe imaginaire en deux points a et b (on supposera que l'ordonnée de b est supérieure à celle de a), séparant C en deux courbes C^- et C^+ (voir la figure 3, $C^- = C \bigcap \{\Re\xi \leq 0\}$, $C^+ = C \bigcap \{\Re\xi \geq 0\}$); les contours fermés Γ^- obtenus en complétant C^- parcouru dans le sens trigonométrique par l'arc $\vec{ab}$, et Γ^+ obtenus en complétant C^+ parcouru dans le même sens par l'arc $\vec{ba}$ sont chacun contenus dans un des demi-plans $\{\Re\xi \leq 0\}$ et $\{\Re\xi \geq 0\}$. Nous avons donc

$$\int_C H(z)dz = \int_{\Gamma^-} H(z)dz + \int_{\Gamma^+} H(z)dz = 0,$$

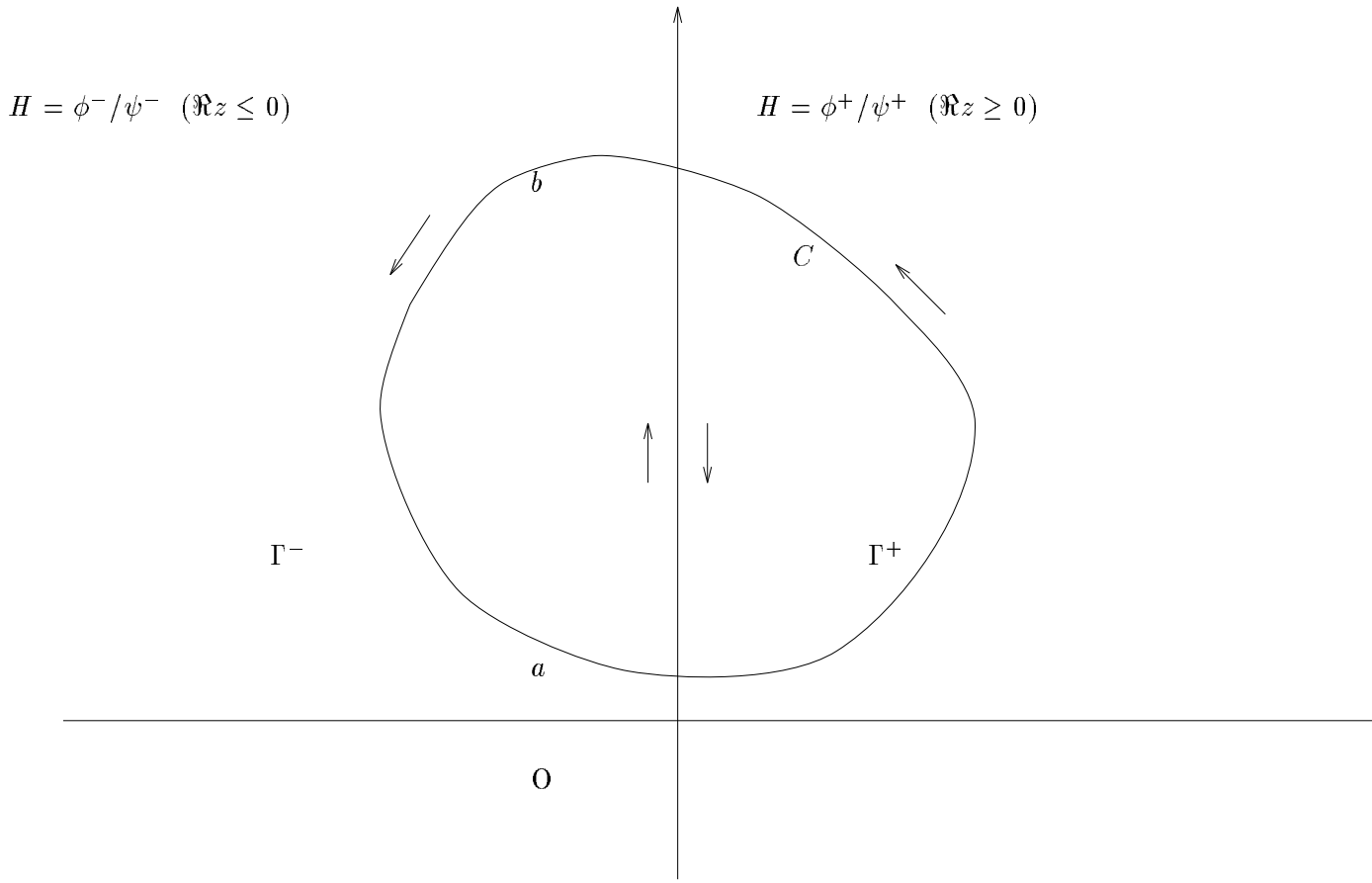

FIGURE 3. La fonction H, et les contours d'intégration C, Γ^- et Γ^+

en appliquant le théorème de Cauchy à la fonction $f(z)$ sur chacun des contours Γ^- et Γ^+.

Par conséquent, sur tout contour fermé C contenu dans $\mathbb{C}$, l'intégrale de $H(z)$ vaut zéro; nous pouvons appliquer le théorème de Morera qui nous dit que $H(z)$ est partout analytique sur $\mathbb{C}$. □

ANNEXE B. PROPRIÉTÉS DE LA LOI DE LA SOMME $Z^+ = S_{\zeta^+}$

B.1. **Continuité.** Nous reprenons les conditions posées pour la démonstration des formules de Karlin-Iglehart. L'incrément X de notre marche aléatoire est tel que $E(X) < 0$, et que $\{P(X > 0) > 0\}$, et d'autre part X est à support compact. Nous effectuons la démonstration dans le cas où la distribution de X est absolument continue, c'est à dire qu'il existe une fonction F telle que $F(x) = \int_{-\infty}^{x} f(t)dt$. Nous supposerons en outre que f est continue par morceaux, ce qui entraîne, puisque X est à support borné, qu'il existe $M < \infty$ tel que, $\{\forall x, F(x + dx) - F(x) < Mdx\}$ (F est Lipschitzienne - voir Gourdon [11] ou Titchmarsh [28]).

Nous noterons $\phi(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx$ la fonction caractéristique de X.

Soit $L_{Z^+}(x) = P(Z^+ < x)$ la loi de Z^+; nous allons démontrer que L_{Z^+}est Lipschitzienne et donc continue. La démonstration utilise un théorème local limite de convergence vers la distribution normale $\mathcal{N}$.

Nous aurons besoin des théorèmes ou lemmes généraux suivants concernant les fonctions caractéristiques.

Théorème 26 (Riemann-Lebesgue). *Si $f(x)$ est intégrable et $\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx$, alors $\phi(t) \to 0$ quand $t \to \pm\infty$.*

Preuve. On peut trouver ce théorème dans Feller [9] □

Lemme 27. *Il existe seulement trois possibilités pour la fonction caractéristique:*

1. $|\phi(t)| < 1$ *pout tout* $t \neq 0$.
2. $|\phi(\tau)| = 1$ *et* $|\phi(t)| < 1$ *pour* $0 < t < \tau$. *Dans ce cas* $|\phi|$ *a période* τ *et il existe un nombre réel* b *tel que* $F(x+b)$ *est arithmétique de pas* $h = 2\pi/\lambda$.
3. $|\phi(t)| = 1$ *pout tout* t. *Dans ce cas* $\phi(t) = e^{ibt}$ *et* F *est concentrée au point* b.

Preuve. (Esquisse.) Nous suivons Feller. Supposons $\phi(\tau) = \int_{\infty}^{\infty} e^{i\tau x} f(x)dx = 1$; dans ce cas l'espérance de la fonction non négative $1 - \cos \tau x$ vaut 0; ceci implique que pour tout intervalle $[a, b]$ ne contenant pas un nombre mh avec $h = \frac{2\pi}{\tau}$, et $m \in \mathbb{Z}$, nous avons $\int_a^b f(x)dx = 0$. La distribution F est donc concentrée aux points mh.

Dans le cas où $|\phi(\tau)| = 1$, il existe b réel tel que $\phi(\tau) = e^{itb}$. La variable $X - b$ a pour fonction caractéristique $\phi(t)e^{-itb}$ qui vaut 1 en $t = \tau$, et le résultat précédant s'applique à la variable translatée. □

Il faut remarquer que cette démonstration ne dit rien sur $\limsup |\phi(t)|$ quand $t \to \infty$.

Durrett [8] (p. 134) donne l'exemple suivant de distribution définie sur $\mathbb{Q}$ qui incite à la prudence en ce qui concerne le comportement limite de la fonction caractéristique:

Exemple 28. *Soit $q_1, q_2, \ldots$ une énumération des rationnels positifs tels que $q_n \leq n$; nous considérons la distribution $P(X = q_n) = P(X = -q_n) = 1/2^{n+1}$; nous avons $E(X) = 0$, $E(X^2) < \sum n^2/2^n < \infty$, mais $\phi(t) = E(e^{itX}) = \sum_{n=1}^{\infty} \cos q_n/2^n$ et $\limsup_{t\to\infty} \phi(t) = 1$.*

Nous allons utiliser également un théorème local limite de convergence (voir Durrett p. 134), de la famille des théorèmes de Berry-Esseen (le théorème de Berry-Esseen est un théorème de vitesse de convergence faible, en loi ou fonction de répartition et non en densité).

Théorème 29. *Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes et identiquement distribuées, avec $EX_i = 0$ et $0 < EX_i^2 = \sigma^2 < \infty$, les X_i ayant une fonction caractéristique commune telle que $|\phi(t)| < 1$ pour tout $t \neq 0$. Soit $S_n = X_1 + \cdots + X_n$ et $\mathbf{n}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$;*

alors, si $x_n/\sqrt{n} \to x$ et $a < b$, on a

$$\sqrt{n} P(S_n \in [x_n + a, x_n + b]) \to (b - a)\mathbf{n}(x)$$

Compte tenu de nos hypothèses sur la variable X, de densité continue par morceaux et à support compact, nous pouvons appliquer le théorème de Riemann-Lebesgue et le lemme 27, qui nous donnent les hypothèses du théorème 29 et en particulier $|\phi(t)| < 1$, $\forall t \neq 0$.

Nous arrivons donc à un théorème de continuité pour la loi de Z^+:

Théorème 30. *Dans les conditions énoncées au début du paragraphe (X est à support compact et a une densité continue par morceaux), la loi L_{Z^+}de Z^+ est Lipschitzienne, et donc continue.*

Preuve. Nous allons étudier la variation de la loi $L_{Z^+}(y)$ pour de petites variations de y. Nous avons

$$P(Z^+ \in [y, y+dy]) \leq \sum_{n=1}^{\infty} P(\zeta^+ = n) P(S_n \in [y, y+dy]). \tag{88}$$

Nous allons utiliser la convergence locale limite de la loi des S_n vers la loi normale. Le théorème 29 s'écrit:

$$\forall \epsilon, \text{ quand } x_n/\sqrt{n} \to x, \exists N, \forall n > N, \quad |\sqrt{n} P(S_n \in [x_n + a, x_n + b]) - (b-a)\mathbf{n}(x)| < \epsilon.$$

Choisissons $\epsilon = \alpha dy$ avec $\alpha < 1$, $a = -2dy$, et $b = 2dy$, et faisons tendre x_n vers $x\sqrt{n} = y$ de sorte que $|x_n - x\sqrt{n}| = |x_n - y| < dy$; nous avons alors

$$x_n + a = x_n - 2dy < y - dy < x_n < y + dy < x_n + 2dy = x_n + b, \quad \text{et}$$

$$\sqrt{n} P(S_n \in [x_n + a, x_n + b]) < (b-a)\mathbf{n}(x) + \epsilon < (b-a)(\frac{1}{\sigma\sqrt{2\pi}} + \alpha) < A_1 dy,$$

avec $A_1 < \infty$. En appliquant le théorème 29, nous avons donc, pour $n > N$

$$\sqrt{n} P(S_n \in [y, y+dy]) < \sqrt{n} P(S_n \in [x_n + a, x_n + b] < A_1 dy, \tag{89}$$

ou encore

$$P(S_n \in [y, y+dy]) < \frac{A_1}{\sqrt{n}} dy, \tag{90}$$

Nous pouvons réécrire l'équation 88, avec N tel que l'équation 90 soit valide, sous la forme

(91)

$$P(Z^+ \in [y, y+dy]) \leq \sum_{n=1}^{N} P(\zeta^+ = n) P(S_n \in [y, y+dy]) + \sum_{n=N}^{\infty} P(\zeta^+ = n) P(S_n \in [y, y+dy])$$

$$< \sum_{n=1}^{N} P(S_n \in [y, y+dy]) + \sum_{n=N}^{\infty} \frac{E(\zeta^+)}{n} P(S_n \in [y, y+dy]), \tag{92}$$

le dernier terme étant obtenu en appliquant l'inégalité de Markov à la variable positive ζ^+.

Nous avons supposé que la densité f de X était continue par morceaux, et à support compact. Par conséquent f est bornée; soit $M < \infty$ bornant $f(x)$.

Soit m le minimum de X; nous pouvons faire la translation $X = m + X'$, et X' est à support compact positif. Soit $S'_n = \sum_{j=1}^{n} X' = S_n + nm$; les X'_j comme les X_j étant indépendants, nous avons

$$\begin{aligned}
P(S_n \in [y, y+dy]) &= P(S'_n = S_n + nm \in [y' = y + nm, y' + dy' = y + nm + dy]) \\
&= dP(S_n < y') \\
&= d\left(\int_{x_1=0}^{y'} f(x_1)dx_1 \int_{x_2=0}^{y'-x_1} f(x_2)dx_2 \cdots \int_{x_n=0}^{y'-x_1\cdots-x_{n-1}} f(x_n)dx_n\right) \\
&< M^n d\left(\int_{x_1=0}^{y'} dx_1 \int_{x_2=0}^{y'-x_1} dx_2 \cdots \int_{x_n=0}^{y'-x_1\cdots-x_{n-1}} dx_n\right) \\
&= M^n d(\frac{y'^n}{n!}) = \frac{M^n y'^{n-1}}{(n-1)!} dy' = \frac{M^n (y+mn)^{n-1}}{(n-1)!} dy;
\end{aligned}$$

nous en déduisons que, pour $n \leq N$, X étant bornée, de même que S_n et S'_n, il existe $A_2 < \infty$ tel que $P(S_n \in [y, y+dy]) < A_2 dy$, pour tout y.

Nous avons donc obtenue une double majoration de la décomposition de l'équation 91, et donc nous pouvons écrire

$$P(Z^+ \in [y, y+dy]) < NA_2 dy + \sum_{n=N}^{\infty} \frac{A_1 E(\zeta^+) dy}{n\sqrt{n}} = A_3 dy \quad \text{avec} \quad A_3 < \infty.$$

Nous en concluons que la loi (au sens de la fonction de répartition) de Z^+ est lipschitzienne, et donc continue. □

B.2. **Intégrabilité Riemannienne directe.** La démonstration de la formule de Karlin-Iglehart nous conduit à rechercher la valeur asymptotique à l'infini d'une fonction inconnue H donnée par une équation du renouveau du type $H = h + H * F$, où $*$ indique l'opération de convolution de deux distributions. Ce résultat se nomme théorème du renouvellement. Nous allons maintenant définir ce qu'est l'intégrabilité Riemannienne directe, qui est utilisée dans les conditions de ce théorème .

Définition 31. *Étant donné un pas δ, soient les sommes de Riemann supérieures et inférieures approximant l'intégrale d'une fonction f définies comme*

$$I^\delta = \sum_{i=0}^{\infty} \delta \ \sup\{f(x) : x \in [i\delta, (i+1)\delta)\} \quad et$$

$$I_\delta = \sum_{i=0}^{\infty} \delta \ \inf\{f(x) : x \in [i\delta, (i+1)\delta)\};$$

la fonction f est alors dite directement Riemann intégrable si I^δ et I_δ approchent la même limite finie quand $\delta \to 0$.

On peut consulter Durrett et Feller pour cette définition rendue nécessaire pour les intégrations sur $[0, \infty]$ par le fait qu'il existe des fonctions continues et Riemann intégrables qui ont des variations fortes, et peuvent même tendre à ne pas être bornées (voir Feller); pour de telles fonctions h, il n'existe pas de solutions à l'équation de renouvellement $H = h + H * U$ dont nous parlerons au prochain paragraphe.

Dans le cas où l'intégration porte sur un intervalle borné, l'intégrabilité Riemannienne directe est équivalente à l'intégrabilité Riemannienne classique, et, en particulier, les fonctions continues par morceaux, qui sont Riemann intégrables, sont aussi directement Riemann intégrables. Le nom d'intégrabilité directe provient du fait que l'on considère directement des sommes sur $[0, +\infty]$, et non la limite d'une intégrale définie sur un segment fini $[0, a]$, quand on fait tendre a vers $+\infty$. Nous avons démontré dans les conditions du théorème 30 que la loi de Z^+ est continue; elle est à support compact, puisque le support de Z^+ est contenu dans celui de X qui a été supposé à support compact. La fonction $z(y) = L_{Z^+}(\infty) - L_{Z^+}(y))e^{\lambda y}$ définie sur $[0, \infty]$ a même support compact que Z^+, et elle est continue comme $L_{Z^+}(y)$. Nous avons donc:

Théorème 32. *Dans les conditions du théorème 30 (X de densité continue par morceaux et à support compact, et $E(\zeta^+) < \infty$), la loi L_{Z^+} de Z^+ et la fonction $z(y) = L_{Z^+}(\infty) - L_{Z^+}(y))e^{\lambda y}$ sont directement Riemann intégrables.*

Le lemme suivant donne une condition nécessaire et suffisante d'intégrabilité Riemannienne directe pour les fonctions positives; nous le donnons pour compléter ce paragraphe sur l'intégrabilité Riemannienne directe.

Lemme 33. *Si $h \geq 0$ est continue, h est directement Riemann intégrable si et seulement si la somme de Riemann supérieure I^δ est bornée pour un δ_0.*

Preuve. Si la fonction h est directement Riemann intégrable, il existe par définition un δ_0 tel que $I_{\delta_0} < \infty$; la condition est donc nécessaire. Elle est suffisante; en effet, supposons que, pour δ_0, nous ayons

$$I^{\delta_0} = \sum_{i=0}^{\infty} \delta \ \sup\{f(x) : x \in [i\delta, (i+1)\delta[\} < \infty.$$

La série donnant I^{δ_0} est bornée et à terme positif; pour tout ϵ, il existe donc un N tel que

$$\sum_{n>N} \delta_0 \ \sup\{h(x); x \in [i\delta_0, (i+1)\delta_0[\} < \epsilon \ ;$$

d'autre part, h étant positive, nous avons

$$\begin{aligned} I^{\delta_0} - I_{\delta_0} &= \sum_{i=0}^{\infty} \{(\sup h(x) - \inf h(x)); \quad x \in [i\delta_0, (i+1)\delta_0[\} \\ &\leq \sum_{i=0}^{N} \{(\sup h(x) - \inf h(x)); \quad x \in [i\delta_0, (i+1)\delta_0[\} \\ &\quad + \sum_{n>N} \delta_0 \ \{\sup h(x) : \quad x \in [i\delta_0, (i+1)\delta_0[\}. \end{aligned} \tag{93}$$

Sur $[0, N\delta_0]$, la fonction h est continue et à support compact; elle est donc uniformément continue; nous pouvons donc trouver un entier $k > 0$ tel que

$$\forall j, \quad 0 < j < N2^k, \quad \{\sup h(x) - \inf h(x); \quad x \in [j\delta_0/2^k, (j+1)\delta_0/2^k]\} < \epsilon;$$

enfin, la dernière somme du terme de droite de l'inégalité 93 décroît quand le pas diminue, et donc quand on passe du pas δ_0 au pas $\delta_0/2^k$; nous avons donc obtenu la majoration

$$I^{\delta_0/2^k} - I_{\delta_0/2^k} \leq N2^k\epsilon + \epsilon;$$

la continuité absolue de h sur $[0, N\delta_0]$ étend ce dernier résultat à tout $\delta < \delta_0/2^k$, ce qui complète notre démonstration. □

Remarque: ce lemme permet de montrer d'une autre manière que L_{Z^+} et la fonction $z(y)$ du théorème 32, fonctions continues et à support compact, sont directement Riemann intégrable.

B.3. **Le résultat asymptotique donné par le théorème du renouvellement.** Dans la section 3, nous avons à résoudre une équation de la forme

$$H = h + H * F, \tag{94}$$

où $*$ représente l'opération de convolution (définie sur le demi-axe positif), F est la distribution d'une variable positive X, h est une fonction connue définie sur l'intervalle $[0, \infty]$, et H est la fonction inconnue. Une telle équation s'appelle équation du renouvellement. La recherche de solution à une telle équation fait partie de la théorie du renouvellement; nous suivons la ligne générale de présentation de Feller et de Durrett, sans donner les démonstrations. On pourra aussi consulter pour la théorie du renouvellement Dacunha-Castelle [5] et Ross [25].

L'équation 94 peut aussi s'écrire

$$H(t) = h(t) + \int_{s=0}^{t} H(t-s)F(ds). \tag{95}$$

Le théorème du renouvellement donne la valeur asymptotique de H à l'infini en fonction de l'intégrale sur $[0, \infty]$ de h et de l'espérance μ de X.

Nous considérons une suite de variables positives indépendantes et identiquement distribuée (X_n) de distribution commune F et de moyenne finie $(E(X_1) = \mu < \infty)$, et des temps de

renouvellement $T_i = X_1 + \cdots + X_i$; (dans un modèle de client-serveur, les T_i peuvent être considérés comme les temps d'arrivée des clients).

Nous allons faire usage de la mesure $U = \sum_{i=1}^{\infty} F^{*i}$, où F^{*i} est la convolée ième de F avec elle-même; cette mesure est appelée mesure du renouvellement. On peut aussi définir de manière équivalente la mesure U sur tout sous-ensemble A de $\mathbb{R}^+$ comme

$$U(A) = \sum_{i=0}^{\infty} P(T_i \in A) \quad \text{avec } T_0 = 0.$$

On obtient une première propriété remarquable en considérant le temps d'arrêt N_t tel que $N_t = \inf\{k : T_k > t\}$: Considérons $U(t) = U([0,t])$; nous avons

$$\{T_i \in [0,t]\} \Longleftrightarrow \{N_t > i\}; \quad \text{et donc}$$

$$U(t) = \sum_{i=0}^{\infty} P(T_i \in [0,t]) = \sum_{i=0}^{\infty} P(N_t > i) = E(N_t).$$

Nous avons alors les deux théorèmes équivalents suivants:

Théorème 34. *(Théorème limite pour le temps d'arrêt). Quand $t \to \infty$, $\frac{N_t}{t} \to \frac{1}{\mu}$ presque sûrement, avec $0 < \mu = EX_i \leq \infty$, et $1/\infty = 0$;*

et, dans les mêmes conditions,

Théorème 35. *Quand $t \to \infty$, $\frac{U(t)}{t} \to \frac{1}{\mu}$.*

Preuve. (En suivant Durrett). On démontre d'abord que $E(N_t) < \infty$, résultat qui nous sera utile par la suite. En effet, nous avons $E(X_i) = \mu > 0$; il existe donc $\delta > 0$ tel que $P(X_i > \delta) = \epsilon > 0$. Choisissons alors K tel que $K\delta \geq t$. Puisque une séquence de K fois X_i consécutifs qui sont supérieurs à δ rendent $T_{mK} > K\delta > t$, et qu'il y a m séquences de K fois X_i dans T_{mK}, nous avons

$$\begin{aligned}
&P(K \text{ consécutifs } X_i \text{ tels que } X_i > \delta) = \epsilon^K,\\
&P(\text{Somme } \Sigma \text{ de } K \text{ consécutifs } X_i \text{ telle que } \Sigma < n\delta)\\
&\qquad < 1 - P(K \text{ consécutifs } X_i \text{ tels que } X_i > \delta) = 1 - \epsilon^K,\\
&P(m \text{ fois } \Sigma < n\delta) < P(\Sigma < n\delta)^m = (1-\epsilon^K)^m.
\end{aligned}$$

Ou encore, plus formellement,

$$\begin{aligned}
&P(X_i > \delta, i = j, \ldots, (j+K)) = \epsilon^K,\\
&P(\sum_{i=j}^{j+K} X_i < n\delta) < 1 - P(X_i > \delta, i = j, \ldots, (j+K)) = 1 - \epsilon^K,\\
&P(\sum_{i=j}^{j+K} X_i < n\delta, \ \ \forall j = l + sK, \ 0 \leq s \leq m) < P(\sum_{i=j}^{j+K} X_i < n\delta)^m = (1-\epsilon^K)^m,
\end{aligned}$$

ce qui nous donne

$$P(N_t > mK) = P(T_{mK} < t) < P(T_{mK} < K\delta) \leq (1-\epsilon^K)^m.$$

Nous en déduisons

$$U(t) = E(N_t) = \sum_{i=0}^{\infty} P(N_t > i) \leq K \sum_{m=0}^{\infty} P(N_t > mK) \leq K \sum_{m=0}^{\infty} (1-\epsilon^K)^m < \infty \tag{96}$$

On utilise ensuite l'équation de Wald (voir Durrett)

$$E(S_{N_t}) = E(X)E(N_t), \quad (N_t \text{ temps d'arrêt}, \ S_n = \sum_{i=1}^{n} X_i), \tag{97}$$

en appliquant cette équation au temps d'arrêt N_t et à la somme T_{N_t} $(\geq t)$; ceci nous donne

$$\mu E(N_t) = \mu U(t) = ET_{N_t} \geq t;$$

$$\text{et, donc} \quad \frac{U(t)}{t} \geq \frac{1}{\mu}. \tag{98}$$

D'autre part, si X est borné ($P(X \leq c) = 1$), étant donné que

$$N_t - 1 < N_t \Rightarrow S_{N_t - 1} < t, \qquad \text{par définition, et que,}$$

$$E(S_{N_t}) = E(S_{N_t-1} + X) = E(S_{N_t-1}) + E(X),$$

on obtient

$$\mu E N_t = \mu U(t) = E(S_{N_t}) \leq t + c; \tag{99}$$

en combinant les équations 98 et 99, on obtient

$$\frac{U(t)}{t} \to \frac{t}{\mu} \quad \text{quand} \quad t \to \infty.$$

On trouvera dans Durrett un complément de démonstration dans le cas où X n'est pas bornée. □

La solution de l'équation de renouvellement 94 s'obtient de manière équivalente comme celle de l'équation

$$H(t) = \int_0^t h(t-s)dU(s); \tag{100}$$

Ceci est immédiat en itérant le développement de H dans l'équation 94.

Le théorème de Blackwell qui suit décrit le comportement à l'infini de la mesure de renouvellement U; son application permettra d'arriver au théorème du renouvellement qui nous donne la valeur asymptotique de H à l'infini.

Théorème 36 (Blackwell). *Si F est non-arithmétique, la mesure de renouvellement U vérifie $U([t, t+h]) \to \frac{h}{\mu}$ quand $t \to \infty$.*

Dans le cas arithmétique, on a:

Théorème 37. *Si F est arithmétique de pas τ, $U([t, t+h]) \to \frac{h}{\mu}$ quand $t \to \infty$ et quand h est un multiple du pas τ.*

On peut consulter Feller dans le cas arithmétique. et, dans le cas continu, Durrett, qui utilise une méthode de couplage avec un processus de renouvellement stationnaire.

Nous avons besoin du lemme suivant concernant le mesure de renouvellement U:

Lemme 38. *Pour tout s et t positifs, on a $U([t, t+s]) \leq U([0, s]) = U(s)$.*

Preuve. Remarquons tout d'abord que N_t est le nombre de renouvellements dans l'espace de temps $[0, t]$, en comptant le renouvellement $T_0 = 0$.

Nous noterons $N_{[t,s]}$ (resp. $N_{]t,s]}$) avec $t < s$ le nombre de renouvellements dans l'espace de temps fermé $[t, s]$ (resp. semi-ouvert $]t, s]$). Intuitivement, l'espérance $U([t, t+s]$ du nombre de renouvellements dans $[t, t+s]$ est au plus $U(t)$, puisque le cas le pire correspond à un renouvellement au temps t.

Plus formellement, nous allons considérer les cas où $\{T_{N_t} = t\}$ (renouvellement en t) et où $\{T_{N_t} \neq t\} = \{T_{N_t} < t\}$ (pas de renouvellement en t). Nous avons

$$E(N_{[t,t+s]}) = E(N_{[t,t+s]}1_{\{T_{N_t}=t\}}) + E(N_{[t,t+s]}1_{\{T_{N_t}<t\}}); \tag{101}$$

Compte tenu de l'indépendance des incréments X_i, nous avons d'une part

$$E(N_{[t,t+s]}1_{\{T_{N_t}=t\}}) = E(N_{[0,s]})P(T_{N_t} = t),$$

et d'autre part

$$\begin{aligned} E(N_{[t,t+s]}1_{\{T_{N_t}\neq t\}}) &= \int_{t^+}^{t+s} P(T_{N_t+1} = u)E((1 + N_{]t+u,t+s]})1_{\{T_{N_t+1}=u\}}1_{\{T_{N_t}\neq t\}})\, du \\ &= \int_{t^+}^{t+s} P(T_{N_t+1} = u)E((1 + N_{]u,t+s]})1_{\{T_{N_t+1}=u\}})P(T_{N_t} < t)\, du, \end{aligned}$$

compte tenu de l'indépendance des événements $\{(1 + N_{]u,t+s]})1_{\{T_{N_t+1}=u\}}; u > t\}$ et $\{T_{N_t} < t\}$.

Mais nous avons aussi

$$E(1 + N_{]t+u,t+s]}1_{\{T_{N_t+1}=u\}}; 0 < u \leq s) = E(N_{[0,s-u]}) \leq E(N_{[0,s]}),$$

puisque l'espérance du nombre de renouvellements dans $[t + u, t + s]$, sachant qu'il y a un renouvellement en $t + u$, est égale au nombre de renouvellement dans $[0, s - u]$, en comptant le renouvellement au temps zéro.

Nous en déduisons que

$$E(N_{[t,t+s]}) \leq E(N_{[0,s]})P(T_{N_t} = t) + E(N_{[0,s]})P(T_{N_t} < t) = E(N_{[0,s]}),$$

ou encore, de manière équivalente, que

$$\forall t > 0,\ \forall s > 0 \qquad U([t, t + s]) \leq U([0, s]). \tag{102}$$

□

Nous arrivons maintenant au théorème du renouvellement lui-même que nous allons démontrer en suivant Durrett.

Théorème 39 (Théorème du renouvellement). *Si H est donné par l'équation de renouvellement $H = h + H * F$, où F est non-arithmétique et h est directement Riemann intégrable, alors la valeur asymptotique de H à l'infini est donnée par*

$$H(t) \to \frac{1}{\mu}\int_0^\infty h(s)ds \qquad \textit{quand} \quad t \to \infty, \tag{103}$$

où μ est la moyenne de la variable de loi F.

(Voir la définition 31 pour l'intégrabilité Riemannienne directe.)

Preuve. Nous allons chercher à résoudre l'équation intégrale $H(t) = \int_0^t h(t-s)dU(s)$ équivalente à l'équation du renouvellement; nous effectuons la démonstration dans le cas général où le support de h est $[0, \infty]$.

Considérons d'abord une fonction h en escalier

$$h(s) = \sum_{k=0}^{\infty} a_k 1_{[k\delta,(k+1)\delta[}(s),$$

où $\sum_{k=0}^{\infty} |a_k| < \infty$. Les équations 96 et 102 nous donnent $U([t, t + \delta]) \leq U([0, \delta]) < \infty$.

Choisissons K tel que

$$\sum_{k\geq K} |a_k| \leq \frac{\epsilon}{2U([0, \delta])};$$

choisissons de plus T tel que, pour $t \geq T$ et $0 \leq k < K$, on ait

$$|a_k|.|U([t-(k+1)\delta, t-k\delta]) - \delta/\mu| \leq \frac{\epsilon}{2K},$$

ce qui possible en raison du théorème de Blackwell ($U([t,t+h]) \to h/\mu$, quand $t \to \infty$).

En considérant $U[a,b] = 0$ quand $a < b \leq 0$ et $U[a,b] = U[0,b]$ quand $a < 0 \leq b$, nous avons

$$\int_0^t h(t-s)dU(s) = \int_0^t h(t)dU(t-s) = \sum_{k=0}^{\infty} a_k U([t-(k+1)\delta, t-k\delta]),$$

et donc

$$\left| \int_0^t h(t-s)dU(s) - \frac{1}{\mu}\sum_{k=0}^{K} a_k\delta \right| < \epsilon/2 + \epsilon/2 = \epsilon,$$

dont nous déduisons en faisant tendre K et t vers l'infini

$$\int_0^t h(t-s)dU(s) \to \frac{1}{\mu}\sum_{k=0}^{\infty} a_k\delta.$$

En remarquant que si h_1 est dominée par h_2, la fonction $H_1(x) = \int_0^t h_1(t-s)dU(s)$ est dominée par la fonction $H_2(x) = \int_0^t h_2(t-s)dU(s)$, nous pouvons appliquer le dernier résultat, pour une fonction h quelconque, et pour un $\delta > 0$ quelconque, aux fonctions en escalier

$$\overline{h}_\delta(x) = \{\sup h(x);\quad x \in [k\delta, (k+1)\delta[\ \} \quad \text{et} \quad \underline{h}_\delta(x) = \{\inf h(x);\quad x \in [k\delta, (k+1)\delta[\ \};$$

ce qui nous donne

$$\begin{aligned}\frac{I_\delta}{\mu} = \lim_{t\to\infty}\int_0^t \underline{h}_\delta(t-s)dU(s) &\leq \liminf_{t\to\infty}\int_0^t h(t-s)dU(s)\\ &\leq \limsup_{t\to\infty}\int_0^t h(t-s)dU(s) \leq \lim_{t\to\infty}\int_0^t \overline{h}_\delta(t-s)dU(s) = \frac{I^\delta}{\mu},\end{aligned}$$

où I_δ et I^δ sont les sommes de Riemann directes supérieures et inférieures définies par la subdivision de pas δ de $[0,\infty]$.

Si h est directement Riemann intégrable, I_δ et I^δ approchent la même limite finie I quand $\delta \to 0$, et nous avons alors

$$H(t) = \int_0^t h(t-s)dU(s) \to \frac{I}{\mu} = \frac{1}{\mu}\int_0^\infty h(s)ds \quad \text{quand} \quad t \to \infty.$$

□

Annexe C. Théorèmes Taubériens

Nous avons utilisé dans le paragraphe 2.3 le théorème Taubérien suivant.

Théorème 40. *Soit une suite de nombres réels $c_n \geq 0$, telle que $\sum_{n=0}^{\infty} c_n r^n$ converge pour $0 \leq r < 1$, alors*

$$\lim r \nearrow 1 \sum_{n=0}^{\infty} c_n r^n = \sum_{n=0}^{\infty} c_n,$$

où $\sum_{n=0}^{\infty} c_n$ peut être fini ou infini.

On peut consulter Titchmarsh [28] ou Gourdon [11] pour les théorèmes Taubériens dans le cas où c_n est une suite de valeurs complexes.

Nous avons besoin d'hypothèses moins fortes (suite réelle positive) pour lesquelles Rudin [26] nous donne le théorème suivant (p. 74) dans le cas où $\sum c_n$ converge.

Théorème 41. *Supposons que, pour la série à termes réels c_n, $\sum c_n$ converge. Soit $f(x) = \sum_{n=0}^{\infty} c_n x^n$. Alors,*

$$lim_{x \to 1} f(x) = \sum c_n.$$

Dans le cas où $\sum c_n = \infty$, en considérant une série c_n à termes positifs, pour tout $A > 0$, il existe un entier N tel que $\sum_{n=0}^{N} c_n > A$.

Nous avons alors $f(x) > \sum_{n=0}^{N} c_n x^n > x^N \sum_{n=0}^{N} c_n > x^N A$, et donc $f(x) \to \infty$ quand $x \to 1$.

Les deux cas, somme finie ou infinie du théorème 40 sont donc traités.

Annexe D. Exemple numérique relatif à la matrice de similarité BLOSUM62

Nous appliquons l'équation 70 au cas de la matrice de similarité entre acides aminés BLOSUM62, qui est donnée au tableau 2.

Le tableau 3 donne la distribution limite des acides aminés pour une requête composée d'une suite d'acides aminés identiques, quand la matrice de scores est la matrice BLOSUM62.

```
Matrix made by matblas from blosum62.iij
BLOSUM
```

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Tableau 2. La matrice BLOSUM62

	A	R	N	D	C	Q	E	G	H	I	L	Pr
A	0.262	0.058	0.044	0.044	0.081	0.058	0.059	0.081	0.042	0.058	0.058	0.081
R	0.042	0.294	0.057	0.031	0.021	0.080	0.057	0.030	0.057	0.021	0.030	0.057
N	0.025	0.045	0.286	0.061	0.016	0.045	0.045	0.045	0.062	0.017	0.017	0.045
D	0.030	0.028	0.073	0.339	0.020	0.054	0.103	0.039	0.039	0.020	0.015	0.054
C	0.015	0.006	0.006	0.006	0.307	0.006	0.004	0.006	0.006	0.011	0.011	0.015
Q	0.029	0.054	0.039	0.039	0.014	0.206	0.075	0.020	0.039	0.015	0.020	0.039
E	0.045	0.061	0.061	0.113	0.016	0.119	0.308	0.032	0.061	0.023	0.023	0.061
G	0.068	0.035	0.068	0.050	0.025	0.035	0.036	0.474	0.035	0.018	0.018	0.068
H	0.012	0.022	0.030	0.016	0.008	0.022	0.022	0.012	0.298	0.008	0.008	0.022
I	0.042	0.021	0.023	0.023	0.041	0.021	0.022	0.016	0.021	0.211	0.110	0.057
L	0.069	0.048	0.037	0.027	0.067	0.048	0.035	0.025	0.035	0.179	0.345	0.093
K	0.042	0.108	0.056	0.041	0.020	0.078	0.077	0.029	0.040	0.021	0.029	0.056
M	0.019	0.018	0.013	0.010	0.018	0.025	0.013	0.009	0.013	0.035	0.048	0.025
F	0.022	0.015	0.016	0.016	0.020	0.015	0.015	0.015	0.029	0.040	0.040	0.040
P	0.037	0.025	0.026	0.036	0.018	0.035	0.035	0.026	0.026	0.018	0.018	0.049
S	0.091	0.049	0.093	0.068	0.049	0.068	0.068	0.068	0.049	0.035	0.035	0.068
T	0.058	0.042	0.058	0.043	0.041	0.042	0.042	0.030	0.030	0.042	0.042	0.058
W	0.005	0.005	0.004	0.004	0.007	0.007	0.005	0.007	0.007	0.005	0.007	0.013
Y	0.018	0.017	0.017	0.013	0.016	0.023	0.017	0.012	0.061	0.023	0.023	0.032
V	0.067	0.025	0.027	0.027	0.048	0.034	0.035	0.025	0.025	0.179	0.093	0.067

	L	K	M	F	P	S	T	W	Y	V	Pr
A	0.058	0.058	0.058	0.042	0.058	0.112	0.081	0.030	0.042	0.081	0.081
R	0.030	0.109	0.041	0.021	0.029	0.041	0.041	0.021	0.030	0.021	0.057
N	0.017	0.045	0.023	0.017	0.023	0.062	0.045	0.012	0.023	0.017	0.045
D	0.015	0.039	0.020	0.020	0.039	0.054	0.039	0.015	0.020	0.020	0.054
C	0.011	0.006	0.011	0.008	0.006	0.011	0.011	0.008	0.008	0.011	0.015
Q	0.020	0.054	0.039	0.014	0.028	0.039	0.028	0.020	0.028	0.020	0.039
E	0.023	0.084	0.032	0.022	0.044	0.061	0.044	0.023	0.032	0.032	0.061
G	0.018	0.035	0.025	0.025	0.035	0.068	0.035	0.035	0.025	0.025	0.068
H	0.008	0.016	0.011	0.016	0.011	0.016	0.011	0.011	0.042	0.008	0.022
I	0.110	0.021	0.079	0.057	0.021	0.030	0.041	0.021	0.041	0.152	0.057
L	0.345	0.049	0.179	0.093	0.034	0.048	0.067	0.048	0.067	0.129	0.093
K	0.029	0.285	0.040	0.021	0.040	0.056	0.040	0.021	0.029	0.029	0.056
M	0.048	0.018	0.128	0.025	0.013	0.018	0.018	0.018	0.018	0.035	0.025
F	0.040	0.015	0.040	0.294	0.011	0.021	0.021	0.056	0.107	0.029	0.040
P	0.018	0.035	0.025	0.013	0.497	0.035	0.035	0.013	0.018	0.025	0.049
S	0.035	0.068	0.049	0.035	0.049	0.253	0.094	0.025	0.035	0.035	0.068
T	0.042	0.042	0.042	0.030	0.042	0.081	0.297	0.030	0.030	0.058	0.058
W	0.007	0.005	0.009	0.018	0.003	0.005	0.007	0.477	0.025	0.005	0.013
Y	0.023	0.017	0.023	0.087	0.012	0.017	0.017	0.062	0.318	0.023	0.032
V	0.093	0.035	0.093	0.048	0.035	0.035	0.067	0.025	0.048	0.249	0.067

TABLEAU 3. Probabilité d'apparition des acides aminés quand la requête est composée d'une suite d'acides aminés identiques. L'entrée m_{ij} est la fréquence d'apparition de l'acide aminé a_j quand la requête est composée uniquement de l'acide aminé a_i. Le champ `Pr` indique la probabilité de fond d'apparition des acides aminés dans les protéines connues. La matrice de scores est la matrice BLOSUM62.

Références

[1] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David Lipman. Basic Local Alignment Search Tool. *J. Mol. Biol.*, 215: 403–410, 1990.

[2] J. Bak and D. J. Newman. *Complex Analysis.* Springer Verlag U.T.M., 1997.

[3] Kai Lai Chung. *A Course in Probability Theory.* Academic Press, 1974. Second edition.

[4] R. V. Churchill. *Complex Variables and Applications.* Mac Graw-Hill, 1960.

[5] D. Dacunha-Castelle and M. Duflo. *Probabilité et Statistiques.* Masson, 1983.

[6] P. Deheuvels. *L'Intégrale.* P.U.F., 1980.

[7] J. Dieudonné. *Éléments d'Analyse.* Gauthier-Villars, 1968.

[8] R. Durrett. *Probability: Theory and Examples.* Duxburry Press, 1996. Second edition.

[9] W. Feller. *An Introduction to Probability theory and Its Applications.* John Wiley, 1966. Second edition.

[10] P. Flajolet and R. Sedgewick. The Average Case Analysis of Algorithms: Complex Asymptotics and Generating Functions. Research Report 2026, Institut National de Recherche en Informatique et en Automatique, September 1993.

[11] X. Gourdon. *Les Maths en tête - Analyse.* Ellipses, 1994.

[12] A. Gut. *Stopped Random Walks: limit theorems and applications.* Applied Probability, 1988.

[13] Donald L. Iglehart. Extremes values in the GI/G/1 queues. *The annals of Mathematical Statistics*, 43(2): 627–635, 1972.

[14] S. Karlin and A. Dembo. Limit distribution of maximal segmental score along Markov-dependant partial sums. *Adv. Appl. Prob.*, 24: 113–140, 1992.

[15] S. Karlin and H.M. Taylor. *A first course in Stochastic Processes.* Academic Press, 1975. Second Edition.

[16] S. Karlin and H.M. Taylor. *A second course in Stochastic Processes.* Academic Press, 1975. Second Edition.

[17] Samuel Karlin and Stephen F. Altschul. Methods for assessing the statistical significance of molecular sequences features by using general scoring schemes. *Proc. Natl. Acad Sci. USA*, 87: 2264–2268, March 1990.

[18] Samuel Karlin and Stephen F. Altschul. Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad Sci. USA*, 90: 5873–5877, June 1993.

[19] Samuel Karlin and Volker Brendel. Chance and statistical significance in protein and DNA sequence analysis. *Science*, 257: 39–49, July 1992.

[20] Samuel Karlin and Amir Dembo. Strong limit theorems of empirical functionals for large excedances of partial sums of i.i.d. variables. *The Annals of Probability*, 19(4): 1737–1755, 1991.

[21] Vo-Khac Khoan. *Distributions, Analyse de Fourier, Opérateurs aux dérivées partielles.* Vuibert, 1972. tome 1.

[22] J. Neveu. Cours de Troisième Cycle. Laboratoire de Probabilités, Université Paris VI, 1983-1984.

[23] Philippe Robert. Cours de DEA - Files d'attente. Laboratoire de Probabilités, Université Paris VI, 1993-1994.

[24] G. Robins, D. E. Wrege, T. Zhang, and W. R. Pearson. On the primer selection problem in polymerase chain reaction experiments. Research Report CS-93-68, Department of Computer Science, University of Virginia, Charlottesville, Virginia 22903, U.S.A., 1993.

[25] S. Ross. *Stochastic Processes.* John Wiley, 1983.

[26] W. Rudin. *Principles of Mathematical Analysis.* Mac Graw-Hill, 1976.

[27] F. Spitzer. *Principles of Random Walk.* Van Nostrand, 1964.

[28] E.C. Titchmarsh. *The theory of functions.* Oxford Science Publications, 1939. Second edition.

[29] E.T. Whittaker and G.T. Watson. *A Course of Modern Analysis.* Cambridge University Press, 1965.

Remerciements. Nous remercions tout particulièrement Philippe Robert dont le cours de DEA nous a largement inspiré, et Sabine Mercier qui a bien voulu relire ce document. Nous remercions également Philippe Dumas pour des discussions éclairantes et plus généralement le projet ALGO pour l'aide technique qu'il nous a apportée.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS
Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
(France)
http: //www.inria.fr
ISSN 0249-6399