Elements related to the largest complete excursion of a reflected BM stopped at a fixed time. Application to local score.

Claudie Chabriac, Agnès Lagnoux, Sabine Mercier, Pierre Vallois

Claudie Chabriac [a] Agnès Lagnoux [a,*] Sabine Mercier [a]
Pierre Vallois [b]

[a] *Institut de Mathématiques de Toulouse, UMR 5219, Université Toulouse 2, 5 Allées Antonio Machado, 31058 Toulouse, France*

[b] *Institut Elie Cartan, Université de Lorraine, CNRS UMR 7502, INRIA, BIGS, Campus Sciences, BP 70239, Vandoeuvre-l'es-Nancy Cedex, 54506, France*

Elements related to the largest complete excursion of a reflected BM stopped at a fixed time. Application to local score.

Résumé

We determine the distribution and the density functions of $\big(U^*(t), \theta^*(t)\big)$, where $U^*(t)$ is the maximum over $[0, g(t)]$ of a reflected Brownian motion (U_s), $\theta^*(t)$ the unique time where this maximum is achieved and $g(t)$ stands for the last zero of (U_s) before t. Let U_n^* and θ_n^* be the analog of $U^*(t)$ and $\theta^*(t)$ where the underlying process is the Lindley process, i.e. the difference between a centered real random walk and its minimum. We prove that $\Big(\frac{U_n^*}{\sqrt{n}}, \frac{\theta_n^*}{n}\Big)$ converges in distribution to $\big(U^*(1), \theta^*(1)\big)$ as $n \to \infty$. This result can be applied to compare long biological sequences.

Key words: Lindley process, local score, Donsker's invariance theorem, reflected Brownian motion, inverse of the local time, Brownian excursions.
MSC : 60 F 17, 60 G 17, 60 G 40, 60 G 44, 60 G 50, 60 G 52, 60 J 55, 60 J 65.

1 Introduction

1.1 Statistical properties of the distribution of the local score is largely used by molecular biologists to extract important features in biological sequences and in particular to determine the most significant one among a collection of biological sequences, see for instance [13] and [18]. The probabilistic model

∗. *Corresponding author. Phone* : +335.61.50.46.11
Email address : lagnoux@univ-tlse2.fr
URL : http ://www.lsp.ups-tlse.fr/Fp/Lagnoux

which is commonly used is the following. Associated with a sequence $(\epsilon_i)_{i\geq 1}$ of independent, centered and reduced random variables, consider

$$S_n = \epsilon_1 + \cdots + \epsilon_n \quad \text{for } n \geq 1 \; ; \quad S_0 = 0. \tag{1.1}$$

$$\underline{S}_n = \min_{0\leq i\leq n} S_i, \quad n \geq 0. \tag{1.2}$$

and

$$U_n = S_n - \underline{S}_n = S_n - \min_{i\leq n} S_i, \quad n \geq 0. \tag{1.3}$$

In biological sequence analysis, (ϵ_i) can for example correspond to the physical or chemical properties of the i-th amino acid or nucleotid of the sequence ; it can also reflect the similarity between components of two sequences. The process (U_n) is non negative and is called the Lindley process (see e.g. ChapIII of [3] for more details and references therein or Chap I [6]). The local score $\overline{U}_n$ is the supremum of the Lindley process up to time n. Molecular biologists are interested by this supremum as it highlights the best part of the studied sequence, the eventual segment of DNA transmitted by a common ancestor for sequence comparison or the best hydrophobic segment of a protein that would thus naturally move in a transmembrane place.

The exact distribution of $\overline{U}_n$ have been determined in [9] using the exponentiation of a suitable matrix and classical tools of Markov chain theory. The given formula in [9] is efficient and accurate for small sequences.

However in practice we often face to long sequences and it is supposed that the sequence (ϵ_n) has a negative trend, which actually means that $\mathbb{E}(\epsilon_n) < 0$. The case $E(\epsilon_n) < 0$ is called the logarithmic case (see [19]) as the local score $\bar{U}_n$ grows with $\ln(n)$ and thus this hypothesis force into focusing on small segment compared to the sequence length. Under this assumption, an asymptotic approximation of the distribution of $\overline{U}_n$ has been established with n growing to infinity (see [13], [11]) using renewal theory. This approximation have given rise to a lot of computational developments and in particular taking the form of the BLAST software which is used by million of biologists everyday ([1], [2], [16]). In the case where $\mathbb{E}(\epsilon_n) = 0$, the authors have determined in [8], the asymptotic behavior of the tail distribution of $\overline{U}_n$ when n goes to infinity and in [12] the rate of convergence is given.

Here, we would like to go further. It is clear that the trajectory of (U_n) can be decomposed in a succession of 0 and excursions above 0. These excursions have an important biological interpretation and in particular the highest one corresponds to the best segment due to the physico chemical property or similarity scores that have been chosen. Note that the local score $\overline{U}_n$ can be viewed as the maximum of the heights of all the excursions up to time n. However, as said above, we are interested here in complete excursions up to a fixed time. This leads us to introduce the maximum U_n^* of the heights of all the *complete* excursions up to time n. The second variable which will play an

important role is θ_n^* the time necessary to reach its maximal height U_n^*. See Section 3 for more details and rigorous definitions of the previous r.v.'s.

We believe that the knowledge of the joint distribution of the pair (U_n^*, θ_n^*) would permit to get more efficient statistical tests than the ones only based on the local score. This point should be developed in a forthcoming paper.

1.2 However, it seems difficult to determine explicitly the law of $(U_n^*,\ \theta_n^*)$ for a fixed n. This difficulty can be overcome considering biological sequences which have a large number of bases. This important feature actually means that in our probabilistic model, it is legitimate to approximate the initial random walk (S_n) by a Brownian motion (B_t) started at 0. See Sections 3.1 and 3.2 for details. Using the functional theorem of convergence of Donsker, the Lindley process (U_k) defined by (1.3) can be compared to

$$\widehat{U}(t) := B(t) - \inf_{0\leq s\leq t} B(s), \quad t \geq 0. \tag{1.4}$$

Recall that one has the following classical identity in law :

$$\Big(|B(t)|,\ t \geq 0\Big) \stackrel{(d)}{=} \Big(B(t) - \min_{0\leq u\leq t} B(u),\ t \geq 0\Big). \tag{1.5}$$

Then, we proceed similarly as for the discrete case, i.e. the setting of Lindley's process. We introduce (see Section 2 for more explicit definitions) :

(1) the local score $\overline{U}(t)$ which is the maximum of the heights of all the excursions of $U(s)$ up to time t,
(2) the maximum $U^*(t)$ of the heights of all the *complete* excursions up to time t,
(3) the time $\theta^*(t)$ taken by $U(s)$ starting from the beginning of the largest excursion to hit the maximal level $U^*(t)$.

The approximation of (U_n) by $(\widehat{U}_t)$ permits to prove that the asymptotic distribution of $\left(\frac{U_n^*}{\sqrt{n}},\ \frac{\theta_n^*}{n}\right)$ as $n \to \infty$ is the one of $\Big(U^*(1), \theta^*(1)\Big)$. Consequently, our initial problem in the discrete setting reduces to determine the joint law of $(U^*(t), \theta^*(t))$, where $t > 0$ is given.

1.3 The law of $\overline{U}(t)$ is known (see either subsection 2.11 in [4] or Lemma 3.2 in [15]). However, at our knowledge, the distributions of $U^*(t)$ and the one of $\Big(U^*(t), \theta^*(t)\Big)$ are unknown. Using the theory of excursions related to the one dimensional Brownian motion, we determine in Theorem 2.2 the density function of the couple $\Big(U^*(t), \theta^*(t)\Big)$. However, as explained at the beginning of the introduction, we are interested in statistical tests based on the joint law of

$(U^*(t), \theta^*(t))$. This actually means that for our purpose, we have to determine quantiles of $(U^*(t), \theta^*(t))$. Unfortunately the expression of the density function is too complicated and does not allow to calculate the distribution function of $(U^*(t), \theta^*(t))$. Let $f :]0, \infty[\times]0, \infty[\to \mathbb{R}$ be a bounded function. In Theorem 2.4 we express the expectation of $f\Big(U^*(t), \theta^*(t)\Big)$ as $E(f(A_1)A_2)$ where the random variable (A_1, A_2) can be simulated more or less easily. Therefore for any $a, b > 0$ the probability $P\Big(U^*(t) \leq a, \theta^*(t) \leq b\Big)$ can be calculated via Monte Carlo simulations. Consequently, our formula gives a theoretical solution of finding quantiles for $(U^*(t), \theta^*(t))$.

In Section 2, we begin with giving theoretical results, i.e. whose related to Brownian motion and in particular the distribution of the pair $\Big(U^*(t), \theta^*(t)\Big)$. We recall in Section 3 the functional approximation of one dimensional Brownian motion by normalized random walks. Then, we deduce in Proposition 3 and Theorem 3.2 that approximation permits to obtain the convergence in distribution of of $\left(\frac{U_n^*}{\sqrt{n}}, \frac{\theta_n^*}{n}\right)$ as $n \to \infty$ towards $\Big(U^*(1), \theta^*(1)\Big)$.

2 Theoretical results

We begin with some notation. $(B(t))$ is a standard Brownian motion started at 0 and $U(t)$ is the reflected Brownian motion, i.e.

$$U(t) := |B(t)|, \quad t \geq 0. \tag{2.6}$$

The excursion (above 0) straddling t starts at $g(t)$ and ends at $d(t)$, namely

$$g(t) = \sup\{s \leq t,\ U(s) = 0\}, \quad d(t) = \inf\{s \geq t,\ U(s) = 0\}. \tag{2.7}$$

Let $\overline{U}(t)$ be the supremum of U over $[0, t]$:

$$\overline{U}(t) := \sup_{0 \leq s \leq t} U(s), \quad t \geq 0. \tag{2.8}$$

Then, the highest height $U^*(t)$ of all the *complete* excursions of the process $(U(r)\ ;\ 0 \leq r \leq t)$ equals

$$U^*(t) := \overline{U}(g(t)) = \sup_{0 \leq s \leq g(t)} U(s), \quad t \geq 0. \tag{2.9}$$

Let $f^*(t)$ be the unique time which achieves the maximum of U over $[0,\ g(t)]$:

$$f^*(t) := \sup\{r \leq g(t)\ ;\ U(r) = U^*(t)\}. \tag{2.10}$$

It is important to introduce the left end-point $g^*(t)$ of the excursion straddling $f^*(t)$:

$$g^*(t) := g(f^*(t)) = \sup\{r \leq f^*(t) \ ; \ U(r) = 0\} \tag{2.11}$$

as well as the right end-point $d^*(t)$ of this excursion :

$$d^*(t) := d(f^*(t)) = \inf\{r \geq f^*(t) \ ; \ U(r) = 0\}. \tag{2.12}$$

It is convenient to visualize the different variables in Figure 2.

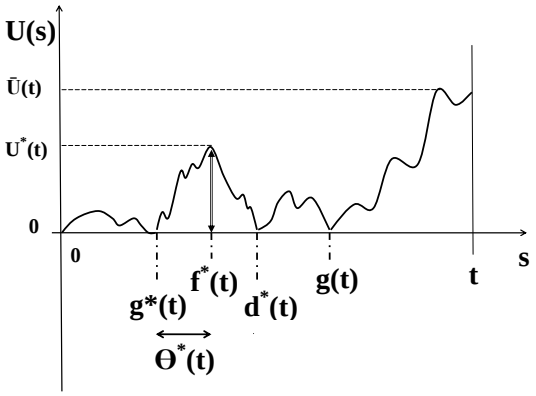

FIGURE 1. Notation $U(t)$

We are interested in the joint law of $U^*(t)$ and $\theta^*(t)$ where the second variable is defined as :

$$\theta^*(t) := f^*(t) - g^*(t), \quad t \geq 0. \tag{2.13}$$

To give either the density function or the expectation of a given function of $\Big(U^*(t), \theta^*(t)\Big)$ it is convenient to introduce preliminary notation.

Notation 2.1 *(1) $(\xi_n)_{n\geq 1} \cup \{\xi, \xi'\}$ is a family of i.i.d. random variables with common density p_ξ such that*

$$\xi \overset{(d)}{=} \xi' \overset{(d)}{=} \xi_n \overset{(d)}{=} T_1(R) \tag{2.14}$$

with

$$T_x(R) = \inf\{s \geq 0 \ ; \ R(s) = x\}, \quad x > 0 \tag{2.15}$$

and $(R(s))$ stands for a 3-dimensional Bessel process started at 0.

(2) $(e_n)_{n\geq 1}$ is a sequence of i.i.d. exponential random variables. We assume moreover that

$$(e_n)_{n\geq 1}, \ (\xi_n)_{n\geq 1}, \ \xi, \ \text{and } \xi' \text{ are independent.} \tag{2.16}$$

(3) Let $(\lambda(x)\ ;\ x \geq 0)$ *be the process :*

$$\lambda(x) := x^2(\xi_1 + \xi_2) + \sum_{k\geq 1} \frac{\xi_{2k+1} + \xi_{2k+2}}{\left(\frac{1}{x} + e_1 + \cdots + e_k\right)^2}, \quad x \geq 0. \tag{2.17}$$

We will prove in Lemma 4.8 below that the previous sum converges a.s. and in L^1*.*

Let (L_t) be the local time process at 0 related to the Brownian motion (B_t). The random function $t \mapsto L_t$ is continuous and non-decreasing. Let $(\tau_s,\ s \geq 0)$ be its right inverse. For more details on the local time process and its inverse, the reader should referred to [14]. In Proposition 1 below we express the distribution of $\left(U^*(t), \theta^*(t)\right)$ in terms of the one of $\left(\overline{U}(\tau_1),\ \tau_1\right)$.

Proposition 1 *Let* t *be a fixed positive real number. Then, for any bounded and Borel function* $f :]0,\infty[\times]0,\infty[\to \mathbb{R}$*,*

$$\mathbb{E}\left[f\left(U^*(t), \theta^*(t)\right)\right] = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \mathbb{E}\left[f(y, y^2\xi)\psi(y) \mathbb{1}_{\{y^2(\xi+\xi')\leq t\}}\right] \frac{dy}{y^2}$$

where

$$\psi(y) := \frac{1}{\tau_1}\left[\sqrt{t - y^2(\xi+\xi')} - \sqrt{\left(t - y^2\left(\xi + \xi' + \frac{\tau_1}{(\overline{U}(\tau_1))^2}\right)\right)_+}\right],$$

$x_+ := \sup\{x, 0\}$ *and* $\xi,\ \xi'$ *have been introduced in Notation 2.1.*

Thus, we are naturally lead to determine the distribution $\left(\overline{U}(\tau_1), \tau_1\right)$.

Proposition 2 *(1) For any* $x \geq 0$*, the sum in (2.17) converges a.s. and in* L^1*.*

(2) The r.v. $\dfrac{1}{\overline{U}(\tau_1)}$ *is exponentially distributed and conditionally on* $\overline{U}(\tau_1) = x > 0$*,*

$$\tau_1 \overset{(d)}{=} \lambda(x). \tag{2.18}$$

Finally, combining Propositions 1 and 2 we get the density function of $\left(U^*(t), \theta^*(t)\right)$.

Theorem 2.2 *For a fixed* $t > 0$*, the couple* $\left(U^*(t), \theta^*(t)\right)$ *has the following density*

$$p_{(U^*(t),\theta^*(t))}(x, y) = \sqrt{\frac{2}{\pi}} \frac{1}{x^4}\, p_\xi\left(\frac{y}{x^2}\right)\ \psi_1\left(x, \frac{y}{x^2}\right)\ \mathbb{1}_{\{x>0, 0<y<t\}} \tag{2.19}$$

with

$$\psi_1(x,u) := \int_0^{+\infty} \mathbb{E}\left[\frac{1}{\lambda(v)}\left\{\sqrt{t-x^2(u+\xi')} - \sqrt{\left(t-x^2\left[u+\xi'+\frac{\lambda(v)}{v^2}\right]\right)_+}\right\}\right.$$
$$\left.\times \mathbb{1}_{\{x^2(u+\xi')\le t\}}\right]\frac{e^{-1/v}}{v^2}\,dv \qquad (2.20)$$

where $x>0$ *et* $0<u<t/x^2$.

Remark 2.3 *The density* p_ξ *is explicitly known, see for instance either [7] or formula 2.0.2 in [5].*

Formula (2.20) has the disadvantage to be not completely explicit and therefore it does not allow a direct calculation of $\mathbb{E}\left[f\Big(U^*(t),\theta^*(t)\Big)\right]$ for a given bounded function f. For instance for our biological motivation it would be interesting to calculate $\mathbb{P}\Big(U^*(t)\le a,\ \theta^*(t)\le b\Big)$ for any $a,b>0$. This leads us to give an equivalent formulation of Theorem 2.2 which gives rise to a more useful formula.

Before stating our second main result, let us introduce e_0, e_0' and G so that :

$$e_0 \text{ and } e_0' \text{ are exponential random variables,} \qquad (2.21)$$
$$G \text{ is a standard Gaussian random variable,} \qquad (2.22)$$
$$e_0, e_0', G \text{ are independent and independent from } \Big\{(\xi_n)_{n\ge1}, \xi, \xi', (e_n)_{n\ge1}\Big\}. \qquad (2.23)$$

Theorem 2.4 *Let* $t>0$. *For any bounded Borel function* $f:\ [0,\infty[\times[0,\infty[\to\mathbb{R}$ *we have :*

$$\mathbb{E}\left[f(U^*(t),\theta^*(t))\right] = \mathbb{E}\left[f\left(\frac{\sqrt{t}|G|}{e_0}, \frac{tG^2\xi}{e_0^2}\right)\frac{e^{\frac{G^2}{2}+e_0}}{\sqrt{1-G^2\left(\frac{\xi+\xi'}{e_0^2}+\lambda\left(\frac{1}{e_0'}\right)\right)}}\right.$$
$$\left.\times \mathbb{1}_{\left\{G^2\left(\frac{\xi+\xi'}{e_0^2}+\lambda\left(\frac{1}{e_0'}\right)\right)\le1,\ e_0<e_0'\right\}}\right] \qquad (2.24)$$

where the r.v.'s e_0, e_0', G, ξ *and* ξ' *are independent and their distributions are given by* (2.21), (2.22) *and* (2.14).

Theorem 2.4 and comes directly by rewritting the proof of Theorem 2.2.

Remark 2.5 *It has been given in [10] an efficient algorithm which permits to simulate very quickly the random variable* ξ. *Although the right hand-side of* (2.24) *is not explicit its forms suggest to develop a Monte Carlo procedure.*

However the difficulty is to simulate the r.v. $\lambda\Big(1/e_0'\Big)$. *Indeed, according to* (2.17) *we have :*

$$\lambda\Big(1/e_0'\Big) = \sum_{k\geq 0} \frac{\xi_{2k+1} + \xi_{2k+2}}{\left(e_0' + e_1 + \cdots + e_k\right)^2}.$$

Therefore to calculate explicitly $\mathbb{P}\Big(U^*(t) \leq a, \theta^*(t) \leq b\Big)$ *demands additional developments.*

The proofs of Theorems 2.2 et 2.4 are postponed in Subsections 4.1 and 4.4.

3 Application to the discrete case

Recall that the r.v. S_n and the Lindley process U_n are associated with the sequence $(\epsilon_i)_{i\geq 1}$ via (1.1) and (1.3) respectively. The process (U_k) is a non negative Markov chain. In the case where $(\epsilon_i)_{i\geq 1}$ are symmetric Bernoulli r.v.'s (i.e. $P(\epsilon_i = \pm 1) = 1/2$), then (U_k) takes its values in $\mathbb{N}$ and moves as a symmetric random walk in $\{1, 2, \cdots\}$ and being at 0, it either stays at this level with probability 1/2 or jumps to 1 with probability 1/2.

In general, the trajectory of (U_k) can be decomposed in a succession of 0 and excursions above 0. An excursion of (U_k) is a process $(e(k),\ 0 \leq k \leq \zeta)$, where

$$e(0) = U(g) = 0, \quad e(\zeta) = U(d) = 0, \quad \zeta := d - g > 0$$

and

$$e(k) := U(g + k) > 0, \text{ for any } \ 0 < k < \zeta.$$

As mentioned in the Introduction, the local score $\overline{U}_n$ is the maximum of (U_k) up to time n and can be interpreted as the maximum of all the heights of the excursions up to time n. Namely :

$$\overline{U}_n := \max_{0\leq k\leq n} U_k, \quad n \geq 0. \tag{3.25}$$

We are interested in the highest *complete* excursion up to time n. We proceed

as in the continuous time setting introducing :

$$
\begin{aligned}
g_n &:= \max\{k \leq n \ ; \ U_k = 0\} \\
U_n^* &:= \overline{U}_{g_n} = \max_{0 \leq k \leq g_n} U_k \\
f_n^* &:= \max\{k \leq g_n; \ U_k = U_n^*\} \\
g_n^* &:= g_{f_n^*} = \max\{k \leq f_n^* \ ; \ U_k = 0\} \\
d_n^* &:= \inf\{k \geq f_n^* \ ; \ U_k = 0\} \\
\theta_n^* &:= f_n^* - g_n^*.
\end{aligned}
\tag{3.26}
$$

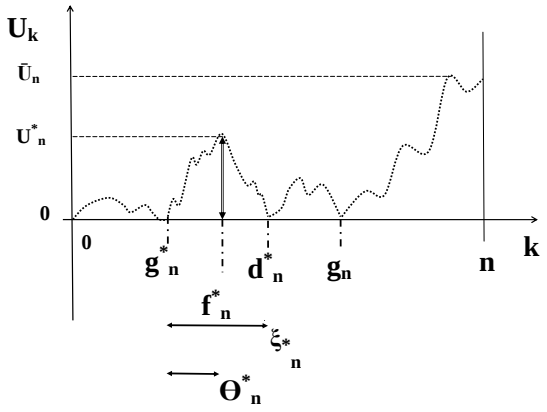

FIGURE 2. Notation U_n

In Section 3.1, we define a continuous process $(U^M(t), t \geq 0)$ as the classical linear interpolation of $(U_n, n \geq 0)$. We naturally introduce the highest high $U^{M,*}(t)$ and length $\theta^{M,*}(t)$ of the complete excursion until time t of $(U^M(t), t \geq 0)$. We conclude linking $(U^{M,*}(1), \theta^{M,*}(1))$ to (U_M^*, θ_M^*). Then we prove in Section 3.2 the convergence of $(U^{M,*}(t), \theta^{M,*}(t))$ to $(U^*(t), \theta^*(t))$. Since the distribution of $(U^*(t), \theta^*(t))$ has been computed in Section 2, we get in such a way an approximation of the distribution of (U_M^*, θ_M^*).

3.1 The linear interpolation of $(U_k)_k$

We keep notation given above and the one introduced in Section 2. Recall in particular that $(B(t),\ t \geq 0)$ stands for a standard Brownian motion started at 0 and $(U(t),\ t \geq 0)$ is the reflected Brownian motion defined by (2.6).

Let $M > 0$ be a scale parameter which permits to obtain the convergence of the normalized random walk to the Brownian motion $(B(t))$ as $M \to \infty$ (see Theorem 3.1). The classical continuous process $\left(B^M(t),\ t \geq 0\right)$ associated with (S_n) and normalizing factor M is classically defined as :

$$B^M\left(\frac{k}{M}\right) = \frac{1}{\sqrt{M}} S_k$$

and

$$B^M(t) = B^M\left(\frac{k}{M}\right) + M\left(t - \frac{k}{M}\right)\left(B^M\left(\frac{k+1}{M}\right) - B^M\left(\frac{k}{M}\right)\right) \quad \frac{k}{M} \leq t \leq \frac{k+1}{M}.$$

We are interested here by the process $(U^M(t),\, t \geq 0)$:

$$U^M(t) = B^M(t) - \min_{s \leq t} B^M(s), \quad t \geq 0. \tag{3.27}$$

Note that

$$U^M\left(\frac{k}{M}\right) = \frac{1}{\sqrt{M}} U_k, \quad k \geq 0 \tag{3.28}$$

where (U_k) is the Lindley process associated with (S_k) via (1.3).

We define the analog of random variables introduced in the discrete setting of Lindley process, see (3.26) and (3.25) in the continous time setting of $\left(B^M(t)\right)$:

$$
\begin{aligned}
\overline{U}^M(t) &:= \sup_{0\le s\le t} U^M(s) \\
g^M(t) &:= \sup\Big\{s \le t\ ;\ U^M(s) = 0\Big\} \\
U^{M,*}(t) &:= \overline{U}^M(g^M(t)) = \sup_{0\le s\le g^M(t)} U^M(s) \\
f^{M,*}(t) &:= \sup\Big\{r \le g^M(t);\ U^M(r) = U^{M,*}(t)\Big\} \\
g^{M,*}(t) &:= g^M(f^{M,*}(t)) = \sup\Big\{r \le f^{M,*}(t)\ ;\ U^M(r) = 0\Big\} \\
d^{M,*}(t) &:= \inf\Big\{s \ge f^{M,*}(t)\ ;\ U(s) = 0\Big\} \\
\theta^{M,*}(t) &:= f^{M,*}(t) - g^{M,*}(t).
\end{aligned}
\tag{3.29}
$$

Using the definition (3.26) of θ^*_M and U^*_M we deduce easily that these r.v.'s can be expressed in terms of their analog in continuous time.

Proposition 3 *We have the following scaling properties :*

$$
\frac{\theta^*_M}{M} = \theta^{M,*}(1) \quad \textit{and} \quad \frac{U^*_M}{\sqrt{M}} = U^{M,*}(1). \tag{3.30}
$$

3.2 Convergence of $\Big(U^{M,*}(t), \theta^{M,*}(t)\Big)$ *to* $\Big(U^*(t), \theta^*(t)\Big)$

Let us start with the Donsker's Theorem, see section 2.10 in [4], which is the key ingredient of our results of convergence.

Theorem 3.1 *The processus* $\Big(B^M(t),\ t \ge 0\Big)$ *converges in distribution to the Brownian motion* $\Big(B(t),\ t \ge 0\Big)$ *when* $M \to +\infty$.

A direct consequence of Theorem 3.1 which be useful for our purpose is the following.

Corollary 4 $\Big(U^M(t),\ t \ge 0\Big) \overset{(d)}{\to} \Big(U(t),\ t \ge 0\Big)$.

Proof Let T be a finite horizon. The set $\mathcal{C}\Big([0,T]\Big)$ of continuous functions defined over $[0,T]$ is equipped with the uniform topology. Since the map $\Lambda : \mathcal{C}\Big([0,T]\Big) \to \mathcal{C}\Big([0,T]\Big)$,

$$\Lambda(\omega)(s) := \omega(s) - \min_{0\le u\le s} \omega(u), \quad s \in [0,T]$$

is continuous, Theorem 3.1 implies that $\Big(U^M(t),\ t \ge 0\Big)$ converges in distribution to $\Big(\widehat{U}(t),\ t \ge 0\Big)$, where $\widehat{U}(t)$ has been already defined in (1.4). Corollary 4 is a direct consequence of the identity in law (1.5). □

Recall that the r.v.'s $g^{M,*}(t)$, $f^{M,*}(t)$, $d^{M,*}(t)$, $\theta^{M,*}(t)$ and $U^{M,*}(t)$ have been defined by relations (3.29). Note that it is unclear that the map $\omega \mapsto \Big(g^{M,*}(t), f^{M,*}(t), d^{M,*}(t), \theta^{M,*}(t)$ is continuous. Therefore the convergence in distribution of $\Big(g^{M,*}(t), f^{M,*}(t), d^{M,*}(t), \theta^{M,*}(t)$, $U^{M,*}(t)\Big)$ as $M \to \infty$ is not a straightforward consequence of Corollary 4.

Theorem 3.2 *Let $t > 0$.*

(1) The 5-uplet $\Big(g^{M,*}(t), f^{M,*}(t), d^{M,*}(t), \theta^{M,*}(t), U^{M,*}(t)\Big)$ converges in distribution to $\Big(g^*(t), f^*(t), d^*(t), \theta^*(t), U^*(t)\Big)$ as $M \to \infty$ where the r.v.'s $g^*(t)$, $f^*(t)$, $d^*(t)$, $\theta^*(t)$, $U^*(t))$ have been defined by relations (2.9)-(2.13).

(2) In particular, $\Big(U^{M,*}(t),\ \theta^{M,*}(t)\Big)$ converges in distribution to $\Big(U^*(t),\ \theta^*(t)\Big)$, as $M \to \infty$.

The proof of Theorem 3.2 is postponed in Appendix 4.5.

4 Proofs

Propositions 1 and 2 are the two main ingredients to prove Theorem 2.2 and we make use of theory of excursions related to Brownian motion $\Big(B(t),\ t \ge 0\Big)$ starting at 0. Recall that $\Big(L_t,\ t \ge 0\Big)$ is the local time process at 0 associated with (B_t) and $\Big(\tau_t,\ t \ge 0\Big)$ its right inverse.

Recall that the random variables $U^*(t)$ and $\theta^*(t)$ have been defined by (2.9) and (2.13).

4.1 Proof of Proposition 1

Lemma 4.1 *Let $f : \mathbb{R}_+^2 \to \mathbb{R}$ a bounded Borel function. Then*

$$\mathbb{E}\Big[f(U^*(t),\theta^*(t))\Big] = \sqrt{\frac{2}{\pi}}\int_0^{+\infty}\mathbb{E}\left[f\Big(\overline{U}(\tau_s),\theta^*(\tau_s)\Big)\frac{1}{\sqrt{t-\tau_s}}\times \mathbb{1}_{\{\tau_s<t\}}\right]ds.$$

Proof The real number $s = L_t$ is the unique s such that $\tau_{s_-} < t < \tau_s$. Thus,

$$f\Big(U^*(t),\theta^*(t)\Big) = \sum_{s\geq 0}\mathbb{1}_{\{\tau_{s_-}<t<\tau_s\}}f\Big(\overline{U}(\tau_{s_-}),\theta^*(\tau_{s_-})\Big)$$

since $B(\tau_{s_-}) = 0$ implies that $U^*(t) = \overline{U}(\tau_{s_-})$ and $\theta^*(t) = \theta^*(\tau_{s_-})$.

Denote e_s the Brownian excursion

$$e_s(v) := \begin{cases} B(\tau_{s_-}+v),\ 0\leq v\leq \tau_s-\tau_{s_-} \text{ for } \tau_s-\tau_{s_-}>0 \\ [\delta] \quad \text{otherwise}\end{cases}$$

and $\zeta(e_s)$ its lifetime : $\zeta(e_s) := \tau_s - \tau_{s_-}$. Since $\tau_s = \tau_{s_-} + \zeta(e_s)$,

$$\mathbb{E}\Big[f(U^*(t),\theta^*(t))\Big] = \mathbb{E}\left[\sum_{s\geq 0} f\Big(\overline{U}(\tau_{s-}),\theta^*(\tau_{s-})\Big)\mathbb{1}_{\{\tau_{s-}<t<\tau_{s-}+\zeta(e_s)\}}\right].$$

Applying Proposition 2.6 deduced from the Master Formula (Proposition 1.10, Chapter XII) in [14], one gets

$$\mathbb{E}\Big[f(U^*(t),\theta^*(t))\Big] = \mathbb{E}\left[\int_0^{+\infty}\left\{\int f\Big(\overline{U}(\tau_s),\theta^*(\tau_s)\Big)\mathbb{1}_{\{\tau_s<t<\tau_s+\zeta(w)\}}\times n(dw)\right\}ds\right]$$

$n(dw)$ being a σ-finite measure on the set of all positive excursions. In particular, by Itô's description of Brownian excursions (Proposition 2.8, Chapter XII in [14])

$$n\Big(\zeta(\omega)>\varepsilon\Big) = \sqrt{\frac{2}{\pi\varepsilon}}$$

from which we deduce

$$\mathbb{E}\Big[f(U^*(t),\theta^*(t))\Big] = \int_0^{+\infty}\mathbb{E}\left[f\Big(\overline{U}(\tau_s),\theta^*(\tau_s)\Big)\mathbb{1}_{\{\tau_s<t\}}\times\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{t-\tau_s}}\right]ds.$$

□

To get the distribution $(U^*(t), \theta^*(t))$, we must determine the one of $(\overline{U}(\tau_s), \theta^*(\tau_s), \tau_s)$. First, we proceed to a change of scale permitting to reduce to the case $s = 1$.

Lemma 4.2 *Let $s > 0$. Then*

$$(\overline{U}(\tau_s), \theta^*(\tau_s), \tau_s) \overset{(d)}{=} (s\overline{U}(\tau_1), s^2\theta^*(\tau_1), s^2\tau_1).$$

Proof According to Corollary 2.2 of [14], τ_s is a stable subordinator of index $1/2$ and as a consequence $\tau_s \overset{(d)}{=} s^2\tau_1$.

Using formula (10) in [17], we have

$$\left(\frac{1}{s}B(s^2r)\ ;\ 0 \le r \le \frac{\tau_s}{s^2}\right) \overset{(d)}{=} (B(r)\ ;\ 0 \le r \le \tau_1) \tag{4.31}$$

Since $B(\tau_s) = 0$,

$$U^*(\tau_s) = \overline{U}(\tau_s) = \max_{r\le\tau_s} U(r) = \max_{r\le\tau_s} |B(r)| = \max_{r'\le\tau_s/s^2} |B(s^2r')| = s \max_{r'\le\tau_s/s^2} \left(\frac{1}{s}B(s^2r')\right).$$

Similarly with $r = s^2r'$,

$$\begin{aligned} f^*(\tau_s) &= \sup\{r < \tau_s\ ,\ |B(r)| = U^*(\tau_s)\} \\ &= \sup\{s^2r' < \tau_s\ ,\ |B(s^2r')| = U^*(\tau_s)\} \\ &= s^2 \sup\left\{r' < \frac{\tau_s}{s^2}\ ,\ \left|\frac{1}{s}B(s^2r')\right| = U^*(\tau_s)\right\} \end{aligned}$$

and

$$g^*(\tau_s) = \sup\{r < f^*(\tau_s)\ ,\ |B(r)| = 0\} = s^2 \sup\left\{r' < \frac{f^*(\tau_s)}{s^2}\ ,\ \left(\frac{1}{s}B(s^2r')\right) = 0\right\}.$$

We conclude using $\theta^*(\tau_s) = f^*(\tau_s) - g^*(\tau_s)$. □

Second, we express the distribution of $\left(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1\right)$ in terms of the one of $\left(\overline{U}(\tau_1), \tau_1\right)$ only.

Lemma 4.3 *Let $h : \mathbb{R}_+^3 \to \mathbb{R}$ be a bounded function. Then*

$$\mathbb{E}\left[h\left(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1\right)\right] = \int_0^{+\infty} \frac{dx}{x^2} \times \mathbb{E}\left[h\Big(x, x^2\xi, x^2(\xi+\xi') + \tau_1\Big)\mathbb{1}_{\{\overline{U}(\tau_1)<x\}}\right].$$

Proof It can be deduced from Theorem 1 of [17] :

$$\mathbb{P}\Big(\overline{U}(\tau_1) < x\Big) = e^{-1/x} \quad \forall\, x > 0 \tag{4.32}$$

and conditionally on $\overline{U}(\tau_1) = x$:

(1) the random variables $f^*(\tau_1) - g^*(\tau_1)$, $d^*(\tau_1) - f^*(\tau_1)$ and $\tau_1 - d^*(\tau_1) + g^*(\tau_1)$ are independent ;

(2) $f^*(\tau_1) - g^*(\tau_1) \overset{(d)}{=} d^*(\tau_1) - f^*(\tau_1) \overset{(d)}{=} T_x(R)$, where $T_x(R)$ has been introduced in Notation 2.1 ;

(3) $\tau_1 - d^*(\tau_1) + g^*(\tau_1)$ is distributed as τ_1 conditionally on $\overline{U}(\tau_1) < x$.

We can see in Figure 4.1 the different variables of interest.

See Remark 4.7 below for a proof of (4.32).

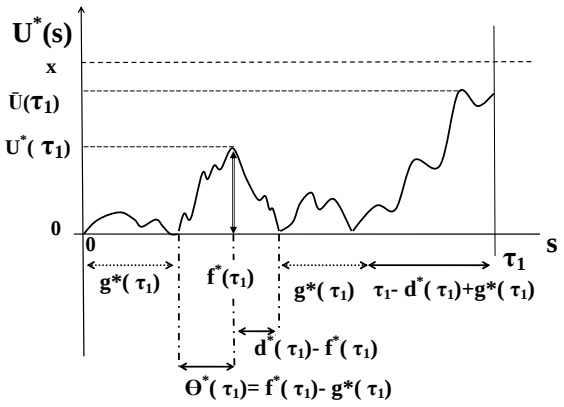

FIGURE 3. Notation with τ_1

For our purpose it is very convenient to introduce a random variable M_x having the same distribution as τ_1 conditionally on $\overline{U}(\tau_1) < x$:

$$\mathbb{E}\Big(\phi(M_x)\Big) = \mathbb{E}\Big(\phi(\tau_1)\Big|\overline{U}(\tau_1) < x\Big)$$

for any bounded Borel function $\phi : [0, \infty[\to \mathbb{R}$.

Assume that $T_x(R)$, $\widetilde{T_x(R)}$, M_x are independent and $\widetilde{T_x(R)}$ is distributed as $T_x(R)$.

Since $\theta^*(\tau_1) = f^*(\tau_1) - g^*(\tau_1)$ and

$$\tau_1 = \Big(\tau_1 - d^*(\tau_1) + g^*(\tau_1)\Big) + \Big(d^*(\tau_1) - f^*(\tau_1)\Big) + \Big(f^*(\tau_1) - g^*(\tau_1)\Big),$$

we deduce

$$\begin{aligned}
\mathbb{E}\Big[h(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1)\Big] &= \int_0^{+\infty} \mathbb{E}\left[h\Big(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1\Big)\Big|\overline{U}(\tau_1) = x)\right] f_{\overline{U}(\tau_1)}(x)dx \\
&= \int_0^{+\infty} \mathbb{E}\left[h\Big(x, T_x(R), T_x(R) + \widetilde{T_x(R)} + M_x\Big)\right] f_{\overline{U}(\tau_1)}(x)dx. \\
&= \int_0^{+\infty} \mathbb{E}\left[h\Big(x, T_x(R), T_x(R) + \widetilde{T_x(R)} + \tau_1\Big)|\overline{U}(\tau_1) < x\right] \frac{e^{-1/x}}{x^2} dx \qquad (4.33)
\end{aligned}$$

The result is a direct consequence of (4.32), (2.14) and the scaling property :

$$T_x(R) \overset{(d)}{=} x^2 T_1(R). \qquad (4.34)$$

□

As a consequence, by Lemmas 4.1, 4.2 and 4.3, we prove Proposition 1.

Proof of Proposition 1

Denote $\Delta := \mathbb{E}\left[f(U^*(t), \theta^*(t))\right]$, where $f : \ [0, \infty[\times[0, \infty[\to \mathbb{R}$ is a bounded Borel function. According to Lemma 4.1, we have :

$$\Delta = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \Delta_s \, ds$$

with

$$\Delta_s := \mathbb{E}\left[f\Big(\overline{U}(\tau_s), \theta^*(\tau_s)\Big) \frac{1}{\sqrt{t - \tau_s}} \mathbb{1}_{\{\tau_s < t\}}\right].$$

Then we use Lemma 4.2

$$\Delta_s = \mathbb{E}\left[f\Big(s\overline{U}(\tau_1), s^2\theta^*(\tau_1)\Big) \frac{1}{\sqrt{t - s^2\tau_1}} \mathbb{1}_{\{s^2\tau_1 < t\}}\right].$$

Finally, Lemma 4.3 leads to

$$\Delta_s = \int_0^{+\infty} \frac{dx}{x^2} \mathbb{E}\left[f(sx, s^2x^2\xi_1) \frac{1}{\sqrt{t - s^2\left[x^2(\xi_1+\xi_2)+\tau_1\right]}} \mathbb{1}_{\{s^2[x^2(\xi_1+\xi_2)+\tau_1]<t, \overline{U}(\tau_1)<x\}} \right]$$
$$= s \int_0^{+\infty} \frac{dy}{y^2} \mathbb{E}\left[f(y, y^2\xi_1) \frac{1}{\sqrt{t - y^2(\xi_1+\xi_2) - s^2\tau_1}} \mathbb{1}_{\{y^2(\xi_1+\xi_2)+s^2\tau_1<t,\ s\overline{U}(\tau_1)<y\}} \right]$$

(recall that, according to Notation 2.1, the r.v.'s ξ_1, ξ_2, τ_1 and $\overline{U}(\tau_1)$ are independent).

We now apply Fubini's theorem

$$\Delta = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \frac{dy}{y^2} \mathbb{E}\left[f(y, y^2\xi_1)\psi(y) \mathbb{1}_{\{y^2(\xi_1+\xi_2)<t\}} \right]$$

where

$$\psi(y) := \int_0^{+\infty} \frac{s\, ds}{\sqrt{t - y^2(\xi_1+\xi_2) - s^2\tau_1}} \mathbb{1}_{\left\{s<\sqrt{\frac{t-y^2(\xi_1+\xi_2)}{\tau_1}} \wedge \left(\frac{y}{\overline{U}(\tau_1)}\right)\right\}}$$
$$= \left[-\frac{1}{\tau_1} \sqrt{t - y^2(\xi_1+\xi_2) - s^2\tau_1} \right]_0^{s_*}$$
$$= \frac{1}{\tau_1} \left[\sqrt{t - y^2(\xi_1+\xi_2)} - \sqrt{t - y^2(\xi_1+\xi_2) - s_*^2\tau_1} \right]$$

with

$$s_* = \frac{y}{\overline{U}(\tau_1)} \wedge \sqrt{\frac{t - y^2(\xi_1+\xi_2)}{\tau_1}}\ ,\ a \wedge b = \inf(a, b).$$

We explicit $\psi(y)$ as :

$$\psi(y) = \begin{cases} \frac{1}{\tau_1}\sqrt{t - y^2(\xi_1+\xi_2)} & \text{if} \quad \frac{t-y^2(\xi_1+\xi_2)}{\tau_1} < \frac{y^2}{(\overline{U}(\tau_1))^2} \\ \frac{1}{\tau_1}\left[\sqrt{t - y^2(\xi_1+\xi_2)} - \sqrt{t - y^2\left(\xi_1+\xi_2+\frac{\tau_1}{(\overline{U}(\tau_1))^2}\right)} \right] & \text{else.} \end{cases}$$

It remains to observe :

$$\psi(y) = \frac{1}{\tau_1} \left[\sqrt{t - y^2(\xi_1+\xi_2)} - \sqrt{\left(t - y^2\left(\xi_1+\xi_2+\frac{\tau_1}{(\overline{U}(\tau_1))^2}\right)\right)_+} \right].$$

□

Remark 4.4 *Since* $\tau \overset{(d)}{=} \frac{1}{G^2}$ *where* G *is a standard Gaussian random variable (see e.g. [14]),*

$$\mathbb{P}(\tau_1 \in ds) = \frac{1}{\sqrt{2\pi}} \frac{1}{s^{3/2}} e^{-\frac{1}{2s}} \mathbb{1}_{[0,+\infty[}(s) ds. \tag{4.35}$$

Consequently $\mathbb{E}\left(\frac{1}{\tau_1}\right) = +\infty$. *Therefore the conditional expectation of the first term in* $\psi(y)$ *is infinite, more precisely,*

$$E\Big(\frac{1}{\tau_1}\sqrt{t - y^2(\xi_1 + \xi_2)}\Big|\xi_1,\ \xi_2\Big) = \infty.$$

4.2 Proof of Proposition 2

As reveals Proposition 1, the law of $\Big(U^*(t),\ \theta^*(t)$ can be expressed in terms of the unknown distribution of the couple $\Big(\tau_1, \overline{U}(\tau_1)\Big)$. The goal of this second step of the proof of Theorem 2.2, is the explicit calculation of this distribution. Since the density function of $\overline{U}(\tau_1)$ is explicit, see (4.32), therefore the distribution of $\Big(\tau_1, \overline{U}(\tau_1)\Big)$ is determined once we know the conditional distribution of τ_1 given $\overline{U}(\tau_1)$. Our proof is based on the study of the process $(\widehat{\lambda}(x),\ x > 0)$ which is supposed to be independent of $(U(t),\ t \geq 0)$ and such that conditionally on $\overline{U}(\tau_1) = x$,

$$\tau_1 \overset{(d)}{=} \widehat{\lambda}(x). \tag{4.36}$$

Obviously (4.36) is equivalent to :

$$E\Big[f(\tau_1) g\Big(\overline{U}(\tau_1)\Big)\Big] = E\Big[f\Big(\widehat{\lambda}\Big(\overline{U}(\tau_1)\Big)\Big) g\Big(\overline{U}(\tau_1)\Big)\Big] \tag{4.37}$$

for any bounded and Borel functions $f,\ g:\ [0,\infty[\to \mathbb{R}$.

We will show that $(\widehat{\lambda}(x))_{x\geq 0}$ satisfies an equation which has a unique solution.

Lemma 4.5 *Let* $x > 0$ *and* $n \geq 1$. *Then,*

$$\widehat{\lambda}(x) \overset{(d)}{=} \Lambda_n + \widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1 + \cdots + e_{n+1}}\right) \tag{4.38}$$

where :

$$\Lambda_n := x^2(\xi_1 + \xi_2) + \sum_{k=1}^{n} \frac{\xi_{2k+1} + \xi_{2k+2}}{\left(\frac{1}{x} + e_1 + \cdots + e_k\right)^2} \quad n \geq 0$$

with the classical convention $\sum_1^0 = 0$.

Proof First we prove for any $x > 0$

$$\widehat{\lambda}(x) \stackrel{(d)}{=} x^2(\xi_1 + \xi_2) + \widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1}\right).$$

Let $f_1,\ f_2 : [0, +\infty[\to [0, +\infty[$ be two bounded and Borel functions and

$$A := \mathbb{E}\left[f_1(\tau_1) f_2\Big((\overline{U}(\tau_1)\Big)\right].$$

Using (4.32) and (4.36) we get :

$$A = \int_0^{+\infty} \frac{1}{x^2} e^{-1/x} f_2(x) \mathbb{E}[f_1(\widehat{\lambda}(x))]\, dx. \tag{4.39}$$

Applying formula (4.33) to $h(x_1, x_2, x_3) = f_1(x_3) f_2(x_1)$ leads to :

$$A = \int_0^{+\infty} \frac{1}{x^2} e^{-1/x} \mathbb{E}\left[f_1\Big(T_x(R) + \widetilde{T_x(R)} + \tau_1\Big) \Big| \overline{U}(\tau_1) < x\right] \times f_2(x)\, dx.$$

Using (4.34), (4.32) (2.14) and (4.37), one gets

$$\begin{aligned} A &= \int_0^{+\infty} \frac{1}{x^2} \mathbb{E}\left[f_1\Big(x^2(\xi_1 + \xi_2) + \tau_1\Big)\, \mathbb{1}_{\{\overline{U}(\tau_1) < x\}}\right] f_2(x)\, dx \\ &= \int_0^{+\infty} \frac{1}{x^2} \left(\int_0^x \frac{e^{-1/y}}{y^2} \mathbb{E}\left[f_1\left(x^2(\xi_1 + \xi_2) + \widehat{\lambda}(y)\right)\right] dy\right) \times f_2(x)\, dx. \end{aligned}$$

Identifying with (4.39) implies :

$$\mathbb{E}\left[f_1\Big(\widehat{\lambda}(x)\Big)\right] = e^{1/x} \int_0^x \frac{e^{-1/y}}{y^2} \mathbb{E}\left[f_1\Big(x^2(\xi_1 + \xi_2) + \widehat{\lambda}(y)\Big)\right] dy.$$

Let Y the random variable defined by $Y = \frac{1}{\frac{1}{x} + e_1}$. Then $0 < Y < x$ and for any $y \in]0, x[$, we get

$$P(Y < y) = P\left(\frac{1}{\frac{1}{x} + e_1} < y\right) = P\left(\frac{1}{x} + e_1 > \frac{1}{y}\right) = P\left(e_1 > \frac{1}{y} - \frac{1}{x}\right) = \exp\left(\frac{1}{x} - \frac{1}{y}\right).$$

Thus the density of Y is $\frac{1}{y^2} e^{1/x} e^{-1/y}\, \mathbb{1}_{[0,x]}(y)$ and

$$\mathbb{E}\left[f_1\Big(\widehat{\lambda}(x)\Big)\right] = \mathbb{E}\left[f_1\Big(x^2(\xi_1 + \xi_2) + \widehat{\lambda}(Y)\Big)\right].$$

The above identity actually means :

$$\widehat{\lambda}(x) \overset{(d)}{=} x^2(\xi_1 + \xi_2) + \widehat{\lambda}(Y), \quad \forall\, x \geq 0.$$

Now we iterate the procedure

$$\widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1}\right) \overset{(d)}{=} \left(\frac{1}{\frac{1}{x} + e_1}\right)^2 (\xi_3 + \xi_4) + \widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1 + e_2}\right)$$

$$\widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1 + e_2}\right) \overset{(d)}{=} \left(\frac{1}{\frac{1}{x} + e_1 + e_2}\right)^2 (\xi_5 + \xi_6) + \widehat{\lambda}\left(\frac{1}{\frac{1}{x} + e_1 + e_2 + e_3}\right)$$

etc...

□

Lemma 4.6 *For any $x > 0$,*

$$\mathbb{E}\left(\widehat{\lambda}(x)\right) = \frac{2}{3}(x + x^2).$$

Proof Using for instance Exercise (4.9) Chap VI in [14]) we get :

$$M_t := \left\{ \cosh\left(\lambda|B(t)|\right) + b \sinh\left(\lambda|B(t)|\right) \right\} \exp\left\{ -\frac{\lambda^2}{2} t - b\lambda L_t \right\}$$

is a local martingale for $\lambda > 0$.

Let $r > 0$, $b = -\dfrac{\cosh(\lambda r)}{\sinh(\lambda r)}$ and

$$\sigma_r := \inf\{s \geq 0\ ;\ |B(s)| = r\} = \inf\{s > 0\ ;\ U(s) = r\}.$$

The process $(M_{t \wedge \tau_1 \wedge \sigma_r}; t \geq 0)$ being bounded, we can apply the stopping theorem, we obtain :

$$\mathbb{E}(M_{\tau_1 \wedge \sigma_r}) = \mathbb{E}(M_0).$$

It is clear that :

$$|B(\sigma_r)| = U(\sigma_r) = r, \quad B(\tau_1) = 0, \quad L_{\tau_1} = 1.$$

Our choice of b implies that $M_{\sigma_r} = 0$. Consequently, $M_{\tau_1 \wedge \sigma_r} = M_{\tau_1} \mathbb{1}_{\{\tau_1 < \sigma_r\}}$ and

$$e^{-b\lambda} \mathbb{E}\left[e^{-\lambda^2 \tau_1/2} \mathbb{1}_{\{\tau_1 < \sigma_r\}}\right] = 1$$

Note that $\{\tau_1 < \sigma_r\} = \{\overline{U}(\tau_1) < r\}$, therefore the above identity can be rewritten as :

$$\mathbb{E}\left[e^{-\lambda^2 \tau_1/2} \mathbb{1}_{\{\overline{U}(\tau_1) < r\}}\right] = e^{b\lambda} = e^{-\lambda \frac{\cosh(\lambda r)}{\sinh(\lambda r)}}. \tag{4.40}$$

Changing λ into $\sqrt{2\lambda}$ in the above identity gives

$$\mathbb{E}\left[e^{-\lambda \tau_1} \mathbb{1}_{\{\overline{U}(\tau_1) < r\}}\right] = \exp\left\{-\sqrt{2\lambda}\, \frac{\cosh(r\sqrt{2\lambda})}{\sinh(r\sqrt{2\lambda})}\right\}. \tag{4.41}$$

Using classical analysis we get :

$$\frac{u \cosh u}{\sinh u} = \frac{u\left(1 + \frac{u^2}{2} + o(u^2)\right)}{u\left(1 + \frac{u^2}{6} + o(u^2)\right)} = \left(1 + \frac{u^2}{2}\right)\left(1 - \frac{u^2}{6}\right) + o(u^2) = 1 + \frac{u^2}{3} + o(u^2) \quad (u \to 0).$$

With $u = r\sqrt{2\lambda}$ and taking the limit $\lambda \to 0$ we obtain :

$$\sqrt{2\lambda}\, \frac{\cosh(r\sqrt{2\lambda})}{\sinh(r\sqrt{2\lambda})} = \frac{1}{r}\left(1 + \frac{2\lambda r^2}{3} + o(\lambda)\right)$$

$$\exp\left\{-\sqrt{2\lambda}\, \frac{\cosh(r\sqrt{2\lambda})}{\sinh(r\sqrt{2\lambda})}\right\} = e^{-1/r} \exp\left\{-\frac{2\lambda r}{3} + o(\lambda)\right\} = e^{-1/r}\left(1 - \frac{2\lambda r}{3} + o(\lambda)\right).$$

Using (4.41) we deduce :

$$\mathbb{E}\left(\tau_1 \mathbb{1}_{\{\overline{U}(\tau_1) < r\}}\right) = \frac{2r}{3}\, e^{-1/r}. \tag{4.42}$$

Let us introduce the following function φ :

$$\varphi(x) := \mathbb{E}\left[\widehat{\lambda}(x)\right] = \mathbb{E}\Big(\tau_1 \Big| \overline{U}(\tau_1) = x\Big), \quad x \geq 0.$$

Therefore :

$$\begin{aligned}
\mathbb{E}\Big(\tau_1 \mathbb{1}_{\{\overline{U}(\tau_1) < r\}}\Big) &= \mathbb{E}\left(\mathbb{E}\Big(\tau_1 | \overline{U}(\tau_1)\Big)\, \mathbb{1}_{\{\overline{U}(\tau_1) < r\}}\right) \\
&= \int_0^r \frac{1}{x^2} e^{-1/x} \mathbb{E}\Big(\widehat{\lambda}(x)\Big)\, dx \\
&= \int_0^r \frac{1}{x^2}\, e^{-1/x} \varphi(x)\, dx.
\end{aligned}$$

Thus formula (4.42) can be rewritten as

$$\int_0^r \frac{1}{x^2}\, e^{-1/x}\varphi(x)\, dx = \frac{2r}{3}\, e^{-1/r} \quad \forall\, r > 0,$$

which conduces to $\varphi(x) = \frac{2}{3}\Big(x + x^2\Big)$.

□

Remark 4.7 *Note that if we take the limit $\lambda \to 0$ in (4.41) we get*

$$\mathbb{P}\Big(\overline{U}(\tau_1) < r\Big) = e^{-1/r}, \quad \forall\, r > 0.$$

Consequently we have given a direct proof of (4.32).

Lemma 4.8 *(1) Λ_n converge a.s. and in L^1 while $n \to +\infty$.*

(2) For any $x > 0$, $\widehat{\lambda}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)$ converges to 0 in L^1, while $n \to +\infty$.

Proof 1) a) As all the random variables in consideration are positive Λ_n converges a.s. as $n \to \infty$ to the limit :

$$\Lambda_\infty := x^2(\xi_1 + \xi_2) + \sum_{k\geq 1} \frac{\xi_{2k+1} + \xi_{2k+2}}{\Big(\frac{1}{x} + e_1 + \cdots + e_k\Big)^2}. \tag{4.43}$$

Λ_∞ is a r.v. which takes its values in $]0, \infty]$. One way to prove that Λ_∞ is a.s. finite is to show that this r.v. has a finite expectation.

b) Recall (cf. Borodin and Salminen [5] p.463)

$$\mathbb{E}\left[e^{-\lambda\xi_1}\right] = \mathbb{E}\left[e^{-\lambda T_1(R)}\right] = \frac{\sqrt{2\lambda}}{\sinh(\sqrt{2\lambda})}, \quad \lambda > 0.$$

An easy calculation leads to

$$\frac{u}{\sinh(u)} = \frac{u}{u + \frac{u^3}{3!} + o(u^3)} = 1 - \frac{u^2}{6} + o(u^2) \quad (u \to 0).$$

Therefore one gets $\mathbb{E}(\xi_1) = \frac{1}{3}$.

c) We are now able to prove that $\mathbb{E}(\Lambda_\infty) < \infty$. Using the fact that $e_1 + \cdots + e_k$ is $\gamma(k)$-distributed, we have successively :

$$\begin{aligned}
\mathbb{E}(\Lambda_\infty) - \frac{2}{3}x^2 &= \mathbb{E}\left(\sum_{k\geq 1} \frac{\xi_{2k+1}+\xi_{2k+2}}{\left(\frac{1}{x}+e_1+\cdots+e_k\right)^2}\right) = \sum_{k\geq 1}\mathbb{E}\left(\frac{\xi_{2k+1}+\xi_{2k+2}}{\left(\frac{1}{x}+e_1+\cdots+e_k\right)^2}\right) \\
&= \sum_{k\geq 1}\Big[\mathbb{E}(\xi_{2k+1})+\mathbb{E}(\xi_{2k+2})\Big]\ \mathbb{E}\left(\frac{1}{\left(\frac{1}{x}+e_1+\cdots+e_k\right)^2}\right) \\
&= \frac{2}{3}\sum_{k\geq 1}\int_0^{+\infty}\frac{1}{\left(\frac{1}{x}+y\right)^2}\frac{y^{k-1}}{(k-1)!}\,e^{-y}\,dy \\
&= \frac{2}{3}\int_0^{+\infty}\frac{1}{\left(\frac{1}{x}+y\right)^2}\left(\sum_{k\geq 1}\frac{y^{k-1}}{(k-1)!}\right)\,e^{-y}\,dy \\
&= \frac{2}{3}\int_0^{+\infty}\frac{dy}{\left(\frac{1}{x}+y\right)^2} < +\infty
\end{aligned}$$

That proves item 1 of Lemma 4.8.

2) We will show that $\widehat{\lambda}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)$ converges to 0 in L^1, $n\to\infty$. Since $\widehat{\lambda}(y)\geq 0$, it is sufficient to check that

$$\lim_{n\to+\infty}\mathbb{E}\left[\widehat{\lambda}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)\right]=0.$$

From Lemma 4.6 we have

$$\mathbb{E}\left[\widehat{\lambda}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)\right]=\frac{2}{3}\left\{\mathbb{E}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)+\mathbb{E}\left(\frac{1}{\left(\frac{1}{x}+e_1+\cdots+e_n\right)^2}\right)\right\}.$$

The Law of Large Numbers implies that

$$\lim_{n\to+\infty}\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}=0\ \text{(a.s.)}$$

Moreover $0<\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\leq x$. Thus Lebesgue's dominated convergence theorem directly implies :

$$\lim_{n\to+\infty}\mathbb{E}\left(\frac{1}{\frac{1}{x}+e_1+\cdots+e_n}\right)=\lim_{n\to+\infty}\mathbb{E}\left(\frac{1}{\left(\frac{1}{x}+e_1+\cdots+e_n\right)^2}\right)=0.$$

□

As a result, it is clear that Proposition 2 is a direct consequence of (4.38) and Lemma 4.8.

4.3 Proof of Theorem 2.2

Using Propositions 1 and 2 we have :

$$\mathbb{E}\left[f(U^*(t),\theta^*(t))\right] = \sqrt{\frac{2}{\pi}}\int_0^{+\infty} \mathbb{E}\left[f(x,x^2\xi)\psi_1(x,\xi)\,\mathbb{1}_{\{x^2\xi\le t\}}\right]\,\frac{dx}{x^2}$$

with

$$\psi_1(x,u) = \mathbb{E}\left[\mathbb{1}_{\{x^2(u+\xi')\le t\}}\int_0^{+\infty}\frac{e^{-1/v}}{v^2}\frac{1}{\lambda(v)}\right.$$
$$\left.\times\left\{\sqrt{t-x^2(u+\xi')}-\sqrt{\left(t-x^2\left(u+\xi'+\frac{\lambda(v)}{v^2}\right)\right)_+}\right\}\,dv\right].$$

As a consequence,

$$\mathbb{E}\left[f(U^*(t),\theta^*(t))\right] = \sqrt{\frac{2}{\pi}}\int_{\mathbb{R}_+^2} f(x,x^2s)\psi_1(x,s)\,\mathbb{1}_{\{x^2s\le t\}}\times p_\xi(s)\frac{dx}{x^2}\,ds.$$

Letting $x^2s = y$ (for x fixed), we deduce that the density of $(U^*(t),\theta^*(t))$ is

$$\sqrt{\frac{2}{\pi}}\psi_1\left(x,\frac{y}{x^2}\right)\,\mathbb{1}_{\{y\le t\}}\,p_\xi\left(\frac{y}{x^2}\right)\frac{1}{x^4}\,\mathbb{1}_{\{x>0,\ 0<y<t\}}.$$

4.4 Proof of Theorem 2.4

The goal is to calculate explicitly $\Delta := \mathbb{E}\left[f(U^*(t),\theta^*(t))\right]$ for particular f. On way is to express Δ as an expectation of a random variable which can be simulated. Then, the Monte Carlo method permits to obtain an approximation of Δ. We revisit the former results of Subsections 4.1 and 4.2. We keep notation given in these subsections.

Interpreting the Lebesgue integral as an expectation in Lemma 4.1 gives :

Lemma 4.9 *Let $t>0$ and f be a bounded function. Then*

$$\mathbb{E}\left[f(U^*(t),\theta^*(t))\right] = \mathbb{E}\left[f\left(\sqrt{t}|G|\overline{U}(\tau_1),tG^2\theta^*(\tau_1)\right)\frac{e^{G^2/2}}{\sqrt{1-G^2\tau_1}}\,\mathbb{1}_{\{G^2\tau_1\le 1\}}\right],$$

G being a standard Gaussian random variable independent of $(U(s)\ ;\ s \geq 0)$.

Proof The density function of $|G|\sqrt{t}$ is $\sqrt{\frac{2}{\pi t}} e^{-s^2/2t} \mathbb{1}_{\{s>0\}}$. Applying Lemma 4.1 we have :

$$\begin{aligned}\mathbb{E}\left[f(U^*(t), \theta^*(t))\right] &= \sqrt{t} \int_0^{+\infty} \sqrt{\frac{2}{\pi t}} e^{-s^2/2t} \left\{ \mathbb{E}\left[f\Big(\overline{U}(\tau_s), \theta^*(\tau_s)\Big) \frac{1}{\sqrt{t-\tau_s}} \mathbb{1}_{\{\tau_s<t\}} \right] e^{s^2/2t} \right\} ds \\ &= \sqrt{t}\, \mathbb{E}\left[f\left(\overline{U}(\tau_{\sqrt{t}|G|}), \theta^*(\tau_{\sqrt{t}|G|})\right) \frac{e^{G^2/2}}{\sqrt{t-\tau_{\sqrt{t}|G|}}} \mathbb{1}_{\{\tau_{\sqrt{t}|G|}\leq t\}} \right].\end{aligned}$$

Lemma 4.9 is a direct consequence Lemma 4.2. □

The analog of Lemma 4.3 is

Lemma 4.10 *Let* $h : [0, \infty[\to \mathbb{R}$ *be a bounded and Borel function. Then*

$$\mathbb{E}\left[h\Big(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1\Big)\right] = \mathbb{E}\left[h\left(\frac{1}{e_0}, \frac{\xi}{e_0^2}, \frac{\xi+\xi'}{e_0^2} + \tau_1\right) \times \mathbb{1}_{\{\overline{U}(\tau_1)e_0<1\}} e^{e_0} \right]$$

where e_0 *is an exponential random variable independent of* $\Big((U(s)\ ;\ s \geq 0), \xi, \xi', G\Big)$.

Proof Let $A := 1/e_0$. Then

$$P(A < x) = P\left(\frac{1}{e_0} < x\right) = P\left(e_0 > \frac{1}{x}\right) = e^{-1/x} \quad \text{and} \quad f_A(x) = \frac{1}{x^2} e^{-1/x}.$$

By Lemma 4.3, one derives

$$\begin{aligned}\mathbb{E}\left[h(\overline{U}(\tau_1), \theta^*(\tau_1), \tau_1)\right] &= \int_0^{+\infty} \frac{1}{x^2}\, e^{-1/x}\, \mathbb{E}\Big[h\Big(x, x^2\xi, x^2(\xi+\xi') + \tau_1\Big)\ \mathbb{1}_{\{\overline{U}(\tau_1)<x\}}\Big] e^{1/x}\, dx \\ &= \mathbb{E}\left[h\Big(A, A^2\xi, A^2(\xi+\xi') + \tau_1\Big) e^{1/A}\, \mathbb{1}_{\{\overline{U}(\tau_1)<A\}}\right].\end{aligned}$$

□

Proof of Theorem 2.4

Synthesizing Lemma 4.9 and 4.10 leads to :

$$\mathbb{E}\left[f(U^*(t),\theta^*(t))\right] = \mathbb{E}\left[f\left(\frac{\sqrt{t}|G|}{e_0}, \frac{tG^2\xi}{e_0^2}\right) \frac{e^{\frac{G^2}{2}+e_0}}{\sqrt{1-G^2\left(\frac{\xi+\xi'}{e_0^2}+\tau_1\right)}} \mathbb{1}_{\left\{G^2\left(\tau_1+\frac{\xi+\xi'}{e_0^2}\right)\leq 1,\ \overline{U}(\tau_1)e_0<1\right\}}\right] \tag{4.44}$$

where f be a bounded and Borel function.

Let e_0' be an exponential random variable independent of e_0, ξ, ξ', σ et $(U(s)\ ;\ s\geq 0)$. It is clear that Proposition 2 is equivalent to the following identity in law :

$$(\overline{U}(\tau_1),\tau_1) \overset{(d)}{=} \left(\frac{1}{e_0'}, \lambda\left(\frac{1}{e_0'}\right)\right). \tag{4.45}$$

Theorem 2.4 is a direct consequence of (4.44) and (4.45). □

4.5 Proof of Theorem 3.2

Recall that the 5-uplet $(g^{M,*}(t), d^{M,*}(t), f^{M,*}(t), \theta^{M,*}(t), U^{M,*}(t))$ has been defined by (3.29). Since $\theta^{M,*}(t) = f^{M,*}(t) - g^{M,*}(t)$, Theorem 3.2 will be proved as soon as the convergence of
$(g^{M,*}(t), d^{M,*}(t), f^{M,*}(t), U^{M,*}(t))$ holds.
Before starting with the proof of the theorem, we go back for a while to the discrete setting.

4.5.1 Auxiliary results in the discrete setting

Let us go back to the random walk defined by (1.1) and introduce for any integer $n_1 > 0$,

$$S_k' := S_{n_1+k} - S_{n_1}, \quad k \geq 0.$$

Lemma 4.11 (Key Property) *(1)* $U_{n+1} = \max(U_n + \epsilon_{n+1}, 0)$.

(2) Let k be an integer such as $k > 0$. Then

$$U_{n_1+i} > 0\ \ \forall i \in \{0,\cdots,k\} \iff U_{n_1} > 0 \text{ and } U_{n_1}+S_i' > 0\ \ \forall i \in \{1,\cdots,k\}.$$

In such a case

$$U_{n_1+i} = U_{n_1} + S_i' \text{ for } 1 \leq i \leq k.$$

Note that

$$U_{n_1} + S_i' > 0 \iff S_{n_1+k} - \underline{S}(0,n_1) > 0. \tag{4.46}$$

Proof

(1) One has

$$\begin{aligned} U_{n+1} &= S_{n+1} - \min_{0\le i\le n+1} S_i = \max_{0\le i\le n+1}(S_{n+1}-S_i) = \max(\max_{0\le i\le n}(S_{n+1}-S_i),0) \\ &= \max(\max_{0\le i\le n}(S_n+\epsilon_{n+1}-S_i),0) = \max(\epsilon_{n+1}+\max_{0\le i\le n}(S_n-S_i),0) \\ &= \max(\epsilon_{n+1}+U_n,0). \end{aligned}$$

(2) First assume $U_{n_1+i} > 0 \ \ \forall i \in \{0,\cdots,k\}$. By 1.,

- $0 < U_{n_1}$.
- $0 < U_{n_1+1} = U_{n_1} + \epsilon_{n_1+1} = U_{n_1} + S'_1$.
- $0 < U_{n_1+2} = U_{n_1+1} + \epsilon_{n_1+2} = U_{n_1} + \epsilon_{n_1+1} + \epsilon_{n_1+2} = U_{n_1} + S'_2 \ \ldots$

Conversely assume $U_{n_1} > 0$ and $U_{n_1} + S'_i > 0 \ \ \forall i \in \{1,\cdots,k\}$. Still by 1.,

- since $U_{n_1} + \epsilon_{n_1+1} > 0$, then $U_{n_1+1} = U_{n_1} + \epsilon_{n_1+1} > 0$.
- since $U_{n_1+1} + \epsilon_{n_1+2} > 0$, then $U_{n_1+2} = U_{n_1+1} + \epsilon_{n_1+2} > 0$.

□

Now consider

$$N := \{g^*_n < n_1,\ n_2 < f^*_n < n_3,\ n_4 < d^*_n < n_5,\ U^*_n \ge b\} \tag{4.47}$$

where $0 < n_1 < \cdots < n_5 < n$ are integers and see (3.26) (resp. (1.3)) for the definition of the r.v.'s $g^*_n, f^*_n, d^*_n, U^*_n$ (resp. (U_k)).
Define

$$\overline{U}(m_1,m_2) := \max_{m_1\le i\le m_2} U_i, \quad \underline{U}(m_1,m_2) := \min_{m_1\le i\le m_2} U_i$$

and

$$n'_i := n_i - n_1, \quad 2 \le i \le 5, \quad n' := n - n_1.$$

The event D can be decomposed as :

$$N = N^1 \cap N^2 \cap N^3 \cap N^4 \tag{4.48}$$

where :

$$N^1 := \Big\{U_k > 0,\ n_1 \le k \le n_4\Big\} = \Big\{\underline{U}(n_1,n_4) > 0\Big\} \tag{4.49}$$

$$N^2 := \Big\{\overline{U}(n_2,n_3) \ge \overline{U}(0,n_2) \vee b\Big\} \tag{4.50}$$

$$N^3 := \Big\{\overline{U}(n_2,n_3) > \overline{U}(n_3,n)\Big\} \tag{4.51}$$

$$N^4 := \Big\{\exists\, k,\ U_k = 0,\ n_4 \le k \le n_5\Big\} = \Big\{\underline{U}(n_4,n_5) \le 0\Big\} \tag{4.52}$$

By the definitions of n_j and (4.46), one has $U_{n_j} > 0$ and $S_{i+n_j} - \underline{S}(0, n_j) > 0 \; \forall i = 1 \dots n_5$. Consequently, we successively have :

$$N^1 = \left\{ \underline{S'}(0, n'_4) > -U_{n_1} \right\} \tag{4.53}$$

$$N^2 = \left\{ \overline{S'}(n'_2, n'_3) \geq -U_{n_1} + \max \left[\overline{U}(0, n_1),\ b,\ \overline{S'}(0, n'_2) + U_{n_1} \right] \right\} \tag{4.54}$$

$$N^3 = \left\{ \overline{S'}(n'_2, n'_3) \geq -U_{n_1} + \overline{U}(n_3, n) \right\} \tag{4.55}$$

$$N^4 = \left\{ \underline{S'}(n'_4, n'_5) \leq -U_{n_1} \right\}. \tag{4.56}$$

The above equalities can be directly read on Figure 4.5.2 (a dash line representing a level that could not be crossed by the process).

We would like to express N^3 with

$$T'_k := S_{t_3+k} - S_{t_3}, \quad k \geq 0.$$

We have

$$\underline{S}(0, n_3 + k) = \min \left\{ \underline{S}(0, n_3),\ S_{n_3} + \underline{T'}(0, k) \right\}$$

and

$$U_{n_3+k} = S_{n_3+k} - \underline{S}(0, n_3 + k) = T'_k + \max \left\{ U_{n_3},\ -\underline{T'}(0, k) \right\}.$$

As a result

$$N^3 = \left\{ \overline{S'}(n'_2, n'_3) > -U_{n_1} + \max_{0 \leq k \leq n - n_3} \left[T'_k + \max \left\{ U_{n_3},\ -\underline{T'}(0, k) \right\} \right] \right\}.$$

4.5.2 Return to the continuous case

Recall that the 5-uplet $(g^{M,*}(t), d^{M,*}(t), f^{M,*}(t), \theta^{M,*}(t), U^{M,*}(t))$ has been defined by (3.29). Since $\theta^{M,*}(t) = f^{M,*}(t) - g^{M,*}(t)$, Theorem 3.2 will be proved as soon as the convergence of
$(g^{M,*}(t), d^{M,*}(t), f^{M,*}(t), U^{M,*}(t))$ holds. We proceed in 5 successive steps.

1) Let $t_1, t_2, \dots, t_5$ be positive real numbers such that $0 < t_1 < \dots < t_5$ and $b > 0$. Let us introduce

$$A^1_M = \{ g^{M,*}(t) < t_1,\ t_2 < f^{M,*}(t) < t_3,\ t_4 < d^{M,*}(t) < t_5,\ U^{M,*}(t) > b \}. \tag{4.57}$$

Consequently the goal is to show

$$\lim_{M \to \infty} P(A^1_M) = P(A^1) \tag{4.58}$$

where

$$A^1 := \{g^*(t) < t_1,\ t_2 < f^*(t) < t_3,\ t_4 < d^*(t) < t_5,\ U^*(t) > b\} \tag{4.59}$$

and the r.v.'s $g^*(t), f^*(t), d^*(t)$ and $U^*(t)$ have been defined by (2.10)-(2.12) and (2.9).

2) In view of the discrete case, let us consider the set of dyadic points

$$D = \bigcup_{m\in\mathbb{N}} D_m \quad \text{where} \quad D_m = \left\{ \frac{k}{2^m}, \quad k \in \{0, 1, \ldots\} \right\}.$$

By density of D in $\mathbb{R}$ and the fact that $D_n \subset D_m$ as soon as $n \leq m$, we can choose without loss of generality strictly positive integers L_0, l and l_i for $i = 1 \ldots 5$ such as

$$t_i = \frac{l_i}{2^{L_0}},\ 1 \leq i \leq 5, \quad t = \frac{l}{2^{L_0}}.$$

Recall also that $\Big(U^M(t),\ t \geq 0\Big)$ is the continuous process defined by (3.27). It is important to note that $U^M(t)$ is also the linear interpolation of $\left(\frac{1}{\sqrt{M}} U_k,\ k \geq 0\right)$.

3) For any continuous function $\omega :\ [0, \infty[\to \mathbb{R}$ denote

$$\overline{\omega}(u, v) := \max_{u\leq r\leq v} \omega(r), \quad \underline{\omega}(u, v) := \min_{u\leq r\leq v} \omega(r), \quad 0 \leq u \leq v. \tag{4.60}$$

Following the procedure presented in the discrete case, the event A^1_M can be decomposed as :

$$A^1_M = A^{1,1}_M \cap A^{1,2}_M \cap A^{1,3}_M \cap A^{1,4}_M \tag{4.61}$$

where :

$$A^{1,1}_M := \Big\{U^M(t) > 0,\ t_1 \leq t \leq t_4\Big\} = \Big\{\underline{U}^M(t_1, t_4) > 0\Big\} \tag{4.62}$$

$$A^{1,2}_M := \Big\{\overline{U}^M(t_2, t_3) \geq \overline{U}^M(0, t_2) \vee b\Big\} \tag{4.63}$$

$$A^{1,3}_M := \Big\{\overline{U}^M(t_2, t_3) > \overline{U}^M(t_3, t)\Big\} \tag{4.64}$$

$$A^{1,4}_M := \Big\{\exists\ s,\ U^M(s) = 0,\ t_4 \leq t \leq t_5\Big\} = \Big\{\underline{U}^M(t_4, t_5) \leq 0\Big\}. \tag{4.65}$$

According to Corollary 4, $\underline{U}^M(t_4, t_5)$ converges in distribution to $\underline{U}(t_4, t_5)$, as $M \to \infty$. Thus we aim at taking the limit when M goes to infinity and applying the following lemma :

Lemma 4.12 *Let (ξ^M) be sequence of $\mathbb{R}^d$-valued r.v.'s which converges in distribution to ξ when $M \to \infty$. Then Porte-Manteau's lemma (see e.g. [4]) asserts that for any Borel Λ in $\mathbb{R}^d$,*

$$\lim_{M\to\infty} \mathbb{P}(\xi^M \in \Lambda) = P(\xi \in \Lambda) \tag{4.66}$$

if $\mathbb{P}(\xi \in \partial\Lambda) = 0$.

Unfortunately, the distribution of $\underline{U}(t_4, t_5)$ (being bounded below by 0) has an atom at 0, therefore we cannot conclude directly that $\lim_{M\to\infty} \mathbb{P}\Big(\underline{U}^M(t_4, t_5) = 0\Big) = \mathbb{P}\Big(\underline{U}(t_4, t_5) = 0\Big)$. This explains why we will introduce the processes W and Z in the following.

4) We follow now the procedure developed in section 4.5.1. It is convenient to introduce :

$$t'_i := t_i - t_1, \quad i \in \{2, 3, 4, 5\}, \quad t' := t - t_1$$

and W^M the process :

$$W^M(s) := B^M(t_1 + s) - B^M(t_1), \quad s \geq 0.$$

The process $\Big(W^M(s), s \geq 0\Big)$ is the linear interpolation of $\Big(\frac{1}{\sqrt{M}}(S_{k+n_1} - S_{n_1}, k \in \mathbb{N}\Big)$.

We deduce from the previous step that $A^1_M = A^2_M$ where :

$$A^2_M := A^{2,1}_M \cap A^{2,2}_M \cap A^{2,3}_M \cap A^{2,4}_M$$

and

$$A^{2,1}_M := \Big\{\underline{W}^M(0, t'_4) > -U^M(t_1)\Big\}$$

$$A^{2,2}_M := \Big\{\overline{W}^M(t'_2, t'_3) \geq -U^M(t_1) + \max\Big[\overline{U}^M(0, t_1),\ b,\ \overline{W}^M(0, t'_2) + U^M(t_1)\Big]\ \Big\}$$

$$A^{2,3}_M := \Big\{\overline{W}^M(t'_2, t'_3) > -U^M(t_1) + \max_{0\leq u\leq t-t_3}\Big[Z^M(u) + \max\Big\{U^M(t_3),\ -\underline{Z}^M(0, u)\Big\}\Big]\Big\}$$

$$A^{2,4}_M := \Big\{\underline{W}^M(t'_4, t'_5) \leq -U^M(t_1)\Big\}.$$

5) To conclude the proof before taking the limit in M, it remains to express

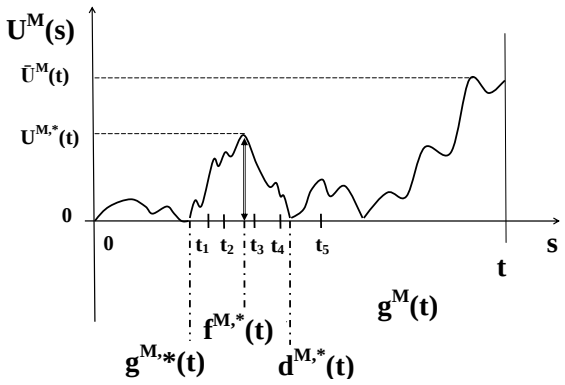

FIGURE 4. Sequence of t_i

the limit subsets in the same way. In that view, let us introduce :

$$W(s) := B(t_1 + s) - B(t_1), \quad Z(s) := B(t_3 + s) - B(t_3), \quad s \geq 0$$

and

$$A^2 := A^{2,1} \cap A^{2,2} \cap A^{2,3} \cap A^{2,4}$$

with

$$A^{2,1} := \left\{ \underline{W}(0, t_4') > -U(t_1) \right\}$$

$$A^{2,2} := \left\{ \overline{W}(t_2', t_4') \geq -U(t_1) + \max \left[\overline{U}(0, t_1),\ b,\ \overline{W}(0, t_2') + U(t_1) \right] \right\}$$

$$A^{2,3} := \left\{ \overline{W}(t_2', t_3') > -U(t_1) + \max_{0 \leq u \leq t - t_3} \left[Z(u) + \max \left\{ U(t_3),\ -\underline{Z}(0, u) \right\} \right] \right\}$$

$$A^{2,4} := \left\{ \underline{W}(t_4', t_5') \leq -U(t_1) \right\}.$$

Recall that for any $u > 0$, the random variables $\max_{0 \leq r \leq u} B(r)$ and $\min_{0 \leq r \leq u} B(r)$ have density. Therefore we can apply (4.66), we get :

$$\lim_{M \to \infty} \mathbb{P}(A^{2,M}) = \mathbb{P}(A^2).$$

Reasoning as in the discrete setting, we get $A^2 = A^1$ where A^1 has been defined by (4.59). It is now clear that (4.58) follows.

■.

Références

[1] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. Basic Local Alignment Search Tool. *JMB*, 215 :403–410, 1990.

[2] S.F. Altschul, T.L. Madden, A.A. Schï¿½ffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. Gapped blast and psi-blast : a new generation of protein database search programs. *Nucleic Acids Res*, 25 :3389–3402, 1997.

[3] Søren Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[4] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[5] Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002.

[6] A. A. Borovkov. *Stochastic processes in queueing theory*. Springer-Verlag, New York, 1976. Translated from the Russian by Kenneth Wickwire, Applications of Mathematics, No. 4.

[7] Z. Ciesielski and S. J. Taylor. First passage times and sojourn times for Brownian motion in space and the exact Hausdorff measure of the sample path. *Trans. Amer. Math. Soc.*, 103 :434–450, 1962.

[8] Etienne M.P. Daudin, J.J. and P. Vallois. Asymptotic behavior of the local score of independent and identically distributed random sequences. *Stoch. Proc. and their Appl.*, 107 :1–28, 2003.

[9] J.J. Daudin and S. Mercier. Exact distribution for the local score of one i.i.d. random sequence. *Jour. Comp. Biol*, 8(4) :373–380, 2001.

[10] M. Deaconu and S. Herrmann. Hitting time for bessel processes - walk on moving spheres algorithm (woms). *The Annals of Applied Probability*, submitted 2011.

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

[12] M.P. Etienne and P. Vallois. Approximation of the supremum of a centered random walk. application to the local score. *Methodology and Computing in Applied Probability*, 6, 2004.

[13] S. Karlin and S.-F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *PNAS*, 87 :2264–2268, 1990.

[14] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental*

Principles of Mathematical Sciences]. Springer-Verlag, Berlin, third edition, 1999.

[15] B. Roynette, P. Vallois, and M. Yor. Penalisations of Brownian motion with its maximum and minimum processes as weak forms of Skorokhod embedding. *Theory Stoch. Process.*, 14(2) :116–138, 2008.

[16] McGinnis S. and Madden T.L. Blast : at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*, 32 :W20–W25, 2004.

[17] P. Vallois. Sur la loi conjointe du maximum et de l'inverse du temps local du mouvement brownien : application à un théorème de Knight. *Stochastics Stochastics Rep.*, 35(3) :175–186, 1991.

[18] M. S. Waterman. *Introduction to Computational Biology : Maps, Sequences and Genomes*. Chapman & Hall, 1995.

[19] M.-S. Waterman, L. Gordon, and R. Arratia. Phase transition in sequence matched and nucleic acid structure. *PNAS*, 84 :1239–1243, 1987.