Bound on the running maximum of a random walk with small drift

Ofer Busani and Timo Seppäläinen

Abstract. We derive a lower bound for the probability that a random walk with i.i.d. increments and small negative drift $\mu$ exceeds the value $x > 0$ by time $N$. When the moment generating functions are bounded in an interval around the origin, this probability can be bounded below by $1 - O(x|\mu|\log N)$. The approach is elementary and does not use strong approximation theorems.

Contents

1. Introduction 52
   1.1. Background 52
   1.2. The question considered 52
   1.3. Sketch of the proof 53
2. Main result 53
3. Comparison with the KMT coupling 56
4. Auxiliary facts 58
5. Proof of the main theorem 59
References 67

Received by the editors July 29th, 2021; accepted September 21st, 2021.

2010 Mathematics Subject Classification. 60G50, 60G17.

Key words and phrases. random walk, running maximum, lower bound.

O. Busani was supported by EPSRC’s EP/R021449/1 Standard Grant.
T. Seppäläinen was partially supported by National Science Foundation grant DMS-1854619 and by the Wisconsin Alumni Research Foundation.
1. Introduction

1.1. Background. This paper arose from the need of a random walk estimate for the authors’ article Busani and Seppäläinen (2020) on directed polymers. This estimate is a positive lower bound on the running maximum of a random walk with a small negative drift. Importantly, the bound had to come with sufficient control over its constants so that it would apply to an infinite sequence of random walks whose drift scales to zero as the maximum is taken over expanding time intervals. The natural approach via a Brownian motion embedding appeared to not give either the desired precision or the uniformity. Hence we resorted to a proof from scratch. For possible wider use we derive the result here under general hypotheses on the distribution of the step of the walk.

The polymer application of the result pertains to the exactly solvable log-gamma polymer on the plane. The objective of Busani and Seppäläinen (2020) is to prove that there are no bi-infinite polymer paths on the planar lattice $\mathbb{Z}^2$. The technical result is that there does not exist any nontrivial Gibbs measures on bi-infinite paths that satisfy the Dobrushin-Lanford-Ruelle (DLR) equations under the Gibbsian specification defined by the quenched polymer measures. In terms of limits of finite polymer distributions, this means that as the southwest and northeast endpoints of a random polymer path are taken to opposite infinities, the middle portion of the path escapes. This is proved by showing that in the limit the probability that the path crosses the $y$-axis along a given edge decays to zero. This probability in turn is controlled by considering stationary polymer processes from the two endpoints to an interval along the $y$-axis. The crossing probability can be controlled in terms of a maximum of a random walk. In the case of the log-gamma polymer, the steps of this random walk are distributed as the difference of two independent log-gamma variables. The case needed for Busani and Seppäläinen (2020) is treated in Example 2.7 below.

1.2. The question considered. We seek a lower bound on the probability that the running maximum of a random walk with negative drift reaches a level $x > 0$. To set the stage, we discuss the matter through Brownian motion. Let $S^N_n = \sum_{i=1}^{n} X^N_i$ be a random walk with drift $E(X^N_1) = \mu_N = \mu N^{-1/2} < 0$, and such that the random walks $S^N_n$ converge weakly to a Brownian motion with drift $\mu < 0$. The probability of the event

$$ \sup_{1 \leq n \leq N} S^N_n > x $$

should be approximately the same as that of

$$ \sup_{0 \leq s \leq 1} (B_s + s\mu) > x N^{-1/2}. $$  \hspace{1cm} (1.1)

This latter can be computed (see (3.7)) to be

$$ \mathbb{P}\{ \sup_{0 \leq s \leq 1} (B_s + s\mu) > x N^{-1/2} \} = 1 - O(|\mu| x N^{-1/2}). $$  \hspace{1cm} (1.2)

This suggests that we should aim for an estimate of the type

$$ \mathbb{P}\left( \sup_{1 \leq n \leq N} S^N_n > x \right) \geq 1 - O(|\mu_N| x). $$ \hspace{1cm} (1.3)

To reach this precision weak convergence is not powerful enough, for a weak approximation of random walk by Brownian motion reaches only a precision of $O(N^{-1/4})$ (Sawyer, 1972; Fraser, 1973). Our estimate (2.4) below does almost capture (1.3): we have to allow an additional $\log N$ factor inside the $O(\cdot)$ and consider $x$ of order at least $(\log N)^2$.

The by-now classical Komlós-Major-Tusnády (KMT) coupling (Komlós et al., 1976) gives a strong approximation of random walk with Brownian motion with a discrepancy that grows logarithmically in time. This precision is sufficient for us, as we illustrate in Section 3. The problem is now the control of the constants in the approximation. Uniformity of the constants is necessary for our application in Busani and Seppäläinen (2020). But verifying this uniformity from the original work
Kolmós et al. (1976) appeared to be a highly nontrivial task. In the end it was more efficient to derive the estimate (Theorem 2.2 below) from the ground up.

The difficulty of the original KMT proof has motivated several recent attempts at simplification and better understanding of the result, such as Bhattacharjee and Goldstein (2016), Chatterjee (2012), and Krishnapur (2020). There is another strong approximation result due to Sakhanenko (1984) which, according to p. 232 of Chatterjee (2012), “is so complex that some researchers are hesitant to use it”.

1.3. Sketch of the proof. Our proof is elementary. The most sophisticated result used is the Berry-Esseen theorem. Given a random walk of small drift $\mu < 0$, our approach can be summarized in two main steps:

(1) Up to the time the random walk hits the level $-\varepsilon|\mu|^{-1}$ it behaves like an unbiased random walk.

(2) By the time the random walk hits the level $-\varepsilon|\mu|^{-1}$ it will have had about $\log_2(\varepsilon|\mu|^{-1}x^{-1})$ independent opportunities to hit the level $x$. By the previous step this implies that the probability on the left-hand side of (1.3) is of order $1 - (1/2)^{\log_2(\varepsilon|\mu|^{-1}x^{-1})} = 1 - O(|\mu|\varepsilon^{-1}x)$. As we will take $\varepsilon = (\log N)^{-1}$ in the proof, we will obtain the right order in (1.3) up to a logarithmic factor (Theorem 2.2).

After the statement of the theorem we illustrate it with examples. Then we demonstrate that even if we knew that the constants in the KMT approximation can be taken uniform, the result would not be essentially stronger in the regime in which we apply our result.

2. Main result

For each $N \in \mathbb{Z}_{>0}$, let $\{X_i^N\}_{i \geq 1}$ be a sequence of non-degenerate i.i.d. random variables. Denote their moment generating function by

$$M_N(\theta) = \mathbb{E}(e^{\theta X_1^N}).$$

Write $M_N^{(0)} = M_N$ and $M_N^{(i+1)} = (d/d\theta)M_N^{(i)}$.

Assumption 2.1. We assume that the random variables $\{X_i^N\}_{i \geq 1}$ satisfy the following:

(i) There exists an open interval $(-\theta_0, \theta_0)$ around the origin on which each moment generating function $M_N$ is finite. Furthermore, there exists a finite constant $C_M$ and $\theta_1 > 0$ such that we have the uniform bounds

$$|M_N^{(i)}(\theta)| \leq C_M \quad \text{for all } N, 0 \leq i \leq 3, \text{ and } \theta \in [-\theta_1, \theta_1]$$

for the compact interval $[-\theta_1, \theta_1] \subset (-\theta_0, \theta_0)$.

(ii) There exists a finite constant $\sigma_* > 0$ such that

$$\mathbb{E}[(X_1^N)^2] \geq \sigma_N^2 = \text{Var}(X_1^N) \geq \sigma_*^2 \quad \text{for all } N.$$  

(2.2)

(iii) There exists a finite constant $D_\mu > 0$ such that the expectations $\mu_N = \mathbb{E}(X_1^N)$ satisfy

$$-D_\mu(\log N)^{-3} \leq \mu_N \leq 0 \quad \text{for all } N.$$  

(2.3)

The conditions in Assumption 2.1 are fairly natural. Note that (2.1) has to be checked only for $i = 0$ at the expense of shrinking the interval $[-\theta_1, \theta_1]$ and increasing $C_M$. To make a positive maximum possible, condition (2.2) ensures enough diffusivity and condition (2.3) limits the strength of the negative drift. The bound (2.4) below shows that $D_\mu$ has to be vanishingly small in order for the result to be nontrivial.

For $m \geq 1$ let $S_m^N = \sum_{k=1}^m X_k^N$ be the random walk associated with the steps $\{X_i^N\}_{i \geq 1}$. Here is the main theorem.
Theorem 2.2. There exist finite constants $C$ and $N_0$ that depend on $\theta_1, \sigma_*^2, D_\mu$ and $C_M$ such that, for every $N \geq N_0$ and $x \geq (\log N)^2$,
\[
P\left( \max_{1 \leq m \leq N} S^N_m \leq x \right) \leq C x(\log N)(|\mu_N| \vee N^{-1/2}).
\] (2.4)

Remark 2.3 (The constants in the theorem). The constant $C$ in the upper bound (2.4) is given by
\[
C = 4 \exp\left\{ 4(C_0 + 4c_\tau^{-1})(1 + D_\mu) + 8\theta_1^{-1}(1 + \log(\theta_1 C_M + 1)) + 4C_M \right\} + 3
\] (2.5)
where
\[
C_0 = 2(\sigma_*^{-1} + 9e^{\theta_1} \sigma_*^{-3} C_M^{5/2}) + 2e^{8C_M \sigma_*^{-4} + 8\sigma_*^{-2} c_\tau^{-1} + 12\sigma_*^{-2}},
\] (2.6)
\[
c_\tau = \log \frac{2}{1 + \Phi_{\sigma_*^2}[-2, 2]},
\] (2.7)
and $\Phi_{\sigma_*^2}$ is the mean zero Gaussian distribution with variance $\sigma_*^2$.

Throughout the proof we state explicitly the various conditions $N \geq N_0$ required along the way. Let us assume that $N \geq 2$ so that $\log N$ does not vanish. Then all the conditions $N \geq N_0$ can be combined into a single condition of the form
\[
f(C_M, D_\mu, \sigma_*^2, \theta_1, N) \geq 1
\] (2.8)
where the function $f$ is a strictly positive continuous function on $\mathbb{R}^4_{>0} \times \mathbb{R}_{\geq 2}$, nondecreasing in $\theta_1$, nonincreasing in $C_M$ and $D_\mu$, but depends on $\sigma_*^2$ in both directions. When $(C_M, D_\mu, \sigma_*^2, \theta_1)$ is restricted to a compact subset $\mathcal{K}$ of $\mathbb{R}^4_{>0}$, there exists a finite index $N_\mathcal{K}$ such that $f(C_M, D_\mu, \sigma_*^2, \theta_1, N)$ is a nondecreasing function of $N \geq N_\mathcal{K}$ for any fixed $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$, and
\[
\lim_{N \to \infty} \inf_{(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}} f(C_M, D_\mu, \sigma_*^2, \theta_1, N) = \infty.
\]

In particular, for each compact subset $\mathcal{K} \subset \mathbb{R}^4_{>0}$ there exists a finite index $N_{\mathcal{K},0}$ such that (2.8) holds for all $N \geq N_{\mathcal{K},0}$ and all $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$. Furthermore, it is evident from (2.5)-(2.7) that $C$ is a continuous function of $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathbb{R}^4_{>0}$. We conclude with the following local uniformity statement.

Corollary 2.4. For each compact subset $\mathcal{K} \subset \mathbb{R}^4_{>0}$ there exist finite constants $C_{\mathcal{K},0}$ and $N_{\mathcal{K},0}$ such that the following holds: the estimate (2.4) with $C = C_{\mathcal{K},0}$ on the right-hand side is valid whenever $N \geq N_{\mathcal{K},0}$, simultaneously for all walks $\{S^N_m\}_{m \geq 1}$ that satisfy Assumption 2.1 with parameters $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$.

We illustrate the result with some examples.

Example 2.5 (Gaussian random walk). Let $B_t$ be a Brownian motion, $\mu < 0$, and define the random walk $S^N_m = B_m + mN^{-1/2}\mu$. We can verify that the bound (2.4) is off by a logarithmic factor in this case, by comparison with the running maximum of the Brownian motion. For $x > 0$ and large enough $N$
\[
P\left( \max_{1 \leq m \leq N} S^N_m \leq x \right) \geq \mathbb{P}\left( \max_{0 \leq t \leq N} B_t + tN^{-1/2}\mu \leq x \right)
\] (2.9)
\[
\geq 1 - e^{-2x N^{-1/2}|\mu|} \geq x|\mu_N| = xN^{-1/2}|\mu|,
\]
where the middle inequality follows from (3.7) with $\mu(N) = \mu$ and $b(N) = xN^{-1/2}$. (2.9) shows that the optimal error is at most $O(x|\mu_N|)$, and that Theorem 2.2, if not optimal, is only log $N$ away from being so.

A natural way to produce examples is to take $X^N_N$ as the difference of two independent random variables whose means come closer as $N$ grows and whose variances stay bounded and bounded away from zero.
Example 2.6 (Exponential walk). Consider a random walk $S_n = \sum_{k=1}^{n} X_k$ with step distribution $X_1^d \sim Y_\alpha - Y_\beta$ where $Y_\alpha$ and $Y_\beta$ are two independent exponential random variables with rates $\alpha$ and $\beta$, respectively. $\mu = \mathbb{E}[X_k] = \frac{1-\beta}{\alpha}$ so we assume that $\alpha > \beta$. The distribution of the supremum of $S_n$ is well-known and also feasible to compute (Example (b) in Section XII.5 of Feller (1971)): for $x > 0$,

$$
P(\sup_{n\geq 0} S_n \leq x) = 1 - \frac{\beta}{\alpha} e^{-(\alpha-\beta)x} = \beta |\mu| (1 + \beta x) + O(\mu^2 x^2)
$$

where we assume $|\mu| x$ small and expand $e^s = 1 + s + O(s^2)$. We obtain a lower bound:

$$
P(\max_{1 \leq n \leq N} S_n \leq x) \geq P(\sup_{n\geq 0} S_n \leq x) = \beta |\mu| (1 + \beta x) + O(\mu^2 x^2)
$$

$$
\geq \beta^2 |\mu| x + O(\mu^2 x^2).
$$

Thus for $|\mu| \geq N^{-1/2}$ and small $x |\mu|$, the upper bound (2.4) loses only a logarithmic factor.

That $\max_{1 \leq n \leq N} S_n$ is close to the overall maximum $\sup_{n \geq 0} S_n$ in the case $|\mu| \geq N^{-1/2}$ is a consequence of the fact that the overall maximum is taken at a random time of order $N^{-2}$. This claim is seen conveniently through ladder intervals $\{T_i\}_{i \geq 1}$. These are the intervals $T_i = \tau_i - \tau_{i-1}$ between successive ladder epochs defined by $\tau_0 = 0$ and

$$
\tau_i = \inf\{n > \tau_{i-1} : S_n > S_{\tau_{i-1}}\}.
$$

The distribution of $T_i$ is given by

$$
P(T_i = \infty) = 1 - \frac{\beta}{\alpha} \quad \text{and} \quad P(T_i = n) = \frac{\alpha^{n-1}\beta^n}{(\alpha + \beta)^{2n-1}} \quad \text{for} \quad n \in \mathbb{Z}_{>0},
$$

where $\{C_n\}_{k \geq 0}$ are the Catalan numbers. (This calculation can be found in Lemma B.3 in the appendix of Fan and Seppäläinen (2020).) Set $T_0 = 0$ and let $N = \max\{n \geq 0 : T_n < \infty\}$ be the number of finite ladder intervals. The maximum $\sup_{n \geq 0} S_n$ is taken at time $\zeta = \sum_{i=1}^{N} T_i$. One calculates $\mathbb{E}[\zeta] = \frac{1}{\alpha+\beta} \mu^{-2}$ and $\mathbb{V}[\zeta] = c_{\alpha,\beta} \mu^{-4}$. Thus for large enough $k$, $P(\zeta > k \mu^{-2}) \leq C_{\alpha,\beta} k^{-2}$.

Example 2.7 (Log-gamma walk). This is the application of Theorem 2.2 used in Busani and Seppäläinen (2020).

Let $G^\lambda$ denote generically a parameter $\lambda$ gamma random variable, that is, $G^\lambda$ has density function $f(x) = \Gamma(\lambda)^{-1} x^{\lambda-1} e^{-x}$ on $\mathbb{R}_{>0}$. For $\alpha, \beta > 0$ let $S_{n, \alpha, \beta} = \sum_{i=1}^{n} X_i^{\alpha, \beta}$ denote the random walk where the distribution of the i.i.d. steps $\{X_i^{\alpha, \beta}\}_{i \geq 1}$ is specified by

$$
X_i^{\alpha, \beta} \overset{d}{=} \log G^\alpha - \log G^\beta
$$

with two independent gamma variables $G^\alpha$ and $G^\beta$ on the right.

Let $\psi_0(s) = \Gamma'(s)/\Gamma(s)$ be the digamma function and $\psi_1(s) = \psi'_0(s)$ the trigamma function on $\mathbb{R}_{>0}$. Their key properties are that $\psi_0$ is strictly increasing with $\psi_0(0^+) = -\infty$ and $\psi_0(\infty) = \infty$, while $\psi_1$ is strictly decreasing and strictly convex with $\psi_1(0^+) = \infty$ and $\psi_1(\infty) = 0$.

Fix a compact interval $[\rho_{\min}, \rho_{\max}] \subset (0, \infty)$. Fix a positive constant $a_0$ and let $\{s_N\}_{N \geq 1}$ be a sequence of nonnegative reals such that $0 \leq s_N \leq a_0 (\log N)^{-3}$. Define a set of admissible pairs

$$
\mathcal{S}_N = \{(\alpha, \beta) : \alpha, \beta \in [\rho_{\min}, \rho_{\max}], -s_N \leq \alpha - \beta \leq 0\}.
$$

For $(\alpha, \beta) \in \mathcal{S}_N$, the mean step satisfies

$$
\mu_{\alpha, \beta} = \mathbb{E}[X_i^{\alpha, \beta}] = \mathbb{E}[\log G^\alpha] - \mathbb{E}[\log G^\beta] = \psi_0(\alpha) - \psi_0(\beta)
$$

$$
= \psi_1(\lambda)(\alpha - \beta) \in [-a_0 \psi_1(\rho_{\min})(\log N)^{-3}, 0]
$$

(2.10)

where we used the mean value theorem with some $\lambda \in (\rho_{\min}, \rho_{\max})$. We take $D_\mu = a_0 \psi_1(\rho_{\min})$. 


The MGF of $X_1^{\alpha,\beta}$ is
\[ M_{\alpha,\beta}(\theta) = \mathbb{E}[e^{\theta X_1}] = \mathbb{E}[(G^\alpha)^\theta] \mathbb{E}[(G^\beta)^{-\theta}] = \frac{\Gamma(\alpha + \theta)\Gamma(\beta - \theta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{2.11} \]
for $\theta \in (-\alpha, \beta)$. For the interval in assumption (2.1) we can take $[-\theta_1, \theta_1] = \left[-\frac{1}{2}\rho_{\min}, \frac{1}{2}\rho_{\min}\right]$. Now (2.1) holds with a single constant $C_M = C_M(\rho_{\min}, \rho_{\max})$ for all choices of $\alpha, \beta \in [\rho_{\min}, \rho_{\max}]$.

The variance satisfies
\[ \text{Var}(X_1^{\alpha,\beta}) = \text{Var}(\log G^\alpha) + \text{Var}(\log G^\beta) = \psi_1(\alpha) + \psi_1(\beta) \geq 2\psi_1(\rho_{\max}) = \sigma_1^2. \]

The constants $(C_M, D_\mu, \sigma_1^2, \theta_1)$ have been fixed and they work simultaneously for all $(\alpha, \beta) \in \mathcal{S}_N$ and all $N \geq 1$. Define $C$ through (2.5)–(2.7). Choose $N_0$ so that (2.8) holds for all $N \geq N_0$. Now $C$ and $N_0$ are entirely determined by $(a_0, \rho_{\min}, \rho_{\max})$. We state the result as a corollary of Theorem 2.2.

**Corollary 2.8.** In the setting described above the bound below holds for all $N \geq N_0$, $(\alpha, \beta) \in \mathcal{S}_N$, and $x \geq (\log N)^2$:
\[ \mathbb{P}\left\{ \max_{1 \leq m \leq N} S_m^{\alpha,\beta} \leq x \right\} \leq C x (\log N)(\mu_{\alpha,\beta} \vee N^{-1/2}). \]

### 3. Comparison with the KMT coupling

As a counterpoint to our Theorem 2.2 we derive here an estimate for a single random walk with the Komlós-Major-Tusnády (KMT) (Komlós et al., 1976) coupling with Brownian motion. We emphasize though that Theorem 3.1 below is not an alternative to our Theorem 2.2 because we do not know how the constants $C, K, \lambda$ below depend on the distribution of the walk. Hence without further work we cannot apply the resulting estimate (3.2) to an infinite family of random walks.

However, this section does illustrate that in a certain regime of vanishing drift the estimates (2.4) and (3.2) are essentially equivalent, as explained below in Remark 3.2. So even if one were to conclude that the constants $C, K, \lambda$ below can be taken uniform, the result remains the same.

Let $\overline{S}_n = \sum_{k=1}^n X_k$ be a mean-zero random walk with i.i.d. steps $\{X_k\}$ and unit variance $\mathbb{E}[X^2] = 1$. The KMT coupling (Theorem 1 in Komlós et al. (1976)) constructs this walk together with a standard Brownian motion $B_\mu$ on a probability space such that the following bound holds:

\[ P\left( \max_{1 \leq k \leq N} |\overline{S}_k - B_k| \geq C \log N + z \right) \leq Ke^{-\lambda z} \quad \text{for all } N \in \mathbb{Z}_{>0} \text{ and } z > 0, \tag{3.1} \]

where $C, K, \lambda$ are finite positive constants determined by the distribution of $X_k$.

We apply this to the running maximum of a random walk with a negative drift.

**Theorem 3.1.** Let $S_n = \sum_{i=1}^n X_i$ be a random walk with i.i.d. steps $\{X_i\}$ that satisfy $\mathbb{E}[e^{tX}] < \infty$ for $t \in (-\delta, \delta)$ for some $\delta > 0$. Assume the drift is negative: $\mu = \mathbb{E}X_1 < 0$, and the variance $\sigma^2 = \mathbb{E}[X_1 - \mu]^2 > 0$. Then there exists a constant $C_1$ determined by the distribution of the normalized variable $\overline{X}_1 = \sigma^{-1}(X_1 - \mu)$ such that, for all real $x > 0$ and integers $N > e^4$,

\[ P\left\{ \max_{0 \leq k \leq N} S_k < x \right\} \leq C_1 \left( N^{1-(\log N)/2} + \frac{\sigma x + \sigma^2 \log N}{N^{3/2}\mu^2} e^{(x+\log N)\sigma^{-1}\mu} \right) \tag{3.2} \]

\[ + 1 - e^{2(\sigma^{-1}x + C_1 \log N)\sigma^{-1}\mu}. \]

**Remark 3.2.** To compare this estimate with Theorem 2.2, imagine that we can let $\mu$ vary as a function of $N$ while preserving the constant $C_1$ in (3.2). Consider the regime where $\sigma^2$ is constant, $x > \log N$ and $|\mu|$ vanishes fast enough so that $x|\mu|$ stays bounded. Then the first parenthetical
expression on the right of (3.2) is dominated by a constant multiple of $xN^{-3/2}\mu^{-2}$. To the last part apply $1 - e^s \leq |s|$ for $s < 0$. The bound (3.2) becomes

$$P\left\{ \max_{0 \leq k \leq N} S_k < x \right\} \leq C x \left( N^{-3/2}\mu^{-2} + |\mu| \right). \quad (3.3)$$

The bound (2.4) is worse than the one above by at most a $\log N$ factor, and not at all if $\mu$ vanishes fast enough. In particular, for the application in Busani and Seppäläinen (2020), the KMT bound cannot give anything substantially better than Theorem 2.2.

**Proof of Theorem 3.1:** Apply (3.1) to the mean-zero unit-variance normalized walk $\overline{S}_N = \sigma^{-1}(S_N - N\mu)$. To simplify some steps below we can assume that $C \geq 1 \lor \lambda^{-1}$. Let $x > 0$ and $z = \lambda^{-1}\log N$.

$$P\left\{ \max_{0 \leq k \leq N} S_k < x \right\} = P\left\{ \max_{0 \leq k \leq N} (\overline{S}_k + k\sigma^{-1}\mu) < \sigma^{-1}x \right\}
\leq Ke^{-\lambda z} + P\left\{ \max_{0 \leq k \leq N} (B_k + k\sigma^{-1}\mu) < \sigma^{-1}x + C \log N + z \right\}. \quad (3.4)$$

Let $M_k = \sup_{0 \leq s < 1} (B_{k+s} - B_k)$. Since $\mu < 0$,

$$\sup_{0 \leq t \leq N} (B_t + t\sigma^{-1}\mu) \leq \max_{0 \leq k \leq N} (B_k + k\sigma^{-1}\mu) + \max_{0 \leq k \leq N-1} M_k.$$  

With this we continue from above.

$$\text{line (3.4)} \leq Ke^{-\lambda z} + P\left\{ \sup_{0 \leq t \leq N} (B_t + t\sigma^{-1}\mu) < \sigma^{-1}x + 2C \log N + z \right\}
+ P\left\{ \max_{0 \leq k \leq N-1} M_k > C \log N \right\}. \quad (3.5)$$

We bound the two probabilities above separately. Recall that $C \geq 1$. For the running maximum of standard Brownian motion, by (2.8.4) on page 96 of Karatzas and Shreve (1991),

$$P\left\{ \max_{0 \leq k \leq N-1} M_k > \log N \right\} \leq NP\left\{ \sup_{0 \leq s \leq 1} B_s > \log N \right\} = N\sqrt{2/\pi} \int_0^\infty e^{-y^2/2} dy
\leq \frac{N\sqrt{2/\pi}}{\log N} \int_0^\infty y e^{-y^2/2} dy = \frac{\sqrt{2/\pi}}{\log N} N^{1 - (\log N)/2}. \quad (3.6)$$

For the running maximum of Brownian motion with drift, use first Brownian scaling, and then the density of the hitting time $T_{b(N)}$ of the point $b(N) = N^{-1/2}(\sigma^{-1}x + 2C \log N + z)$ with drift $\mu(N) = \sigma^{-1}N^{1/2}\mu < 0$ from (3.5.12–3.5.13) on page 197 of Karatzas and Shreve (1991).

$$P\left\{ \sup_{0 \leq t \leq N} (B_t + t\sigma^{-1}\mu) < \sigma^{-1}x + 2C \log N + z \right\}
= P\left\{ \sup_{0 \leq t \leq N} (B_t + t\sigma^{-1}N^{1/2}\mu) < N^{-1/2}(\sigma^{-1}x + 2C \log N + z) \right\}
= P\left\{ \sup_{0 \leq t \leq 1} (B_t + t\mu(N)) < b(N) \right\} = P^{(\mu(N))}(T_{b(N)} > 1)
= b(N) \int_1^\infty \frac{1}{\sqrt{2\pi s^3}} e^{-(b(N) - \mu(N)s)^2/2s} ds + P^{(\mu(N))}(T_{b(N)} = \infty)
= b(N) e^{b(N)\mu(N)} \int_1^\infty \frac{1}{\sqrt{2\pi s^3}} e^{-\frac{1}{2}b(N)^2s^{-1} - \frac{1}{2}\mu(N)^2s} ds + 1 - e^{2b(N)\mu(N)}
\leq 2e^{b(N)\mu(N)} \frac{b(N)}{\mu(N)^2} + 1 - e^{2b(N)\mu(N)}
\leq 2e^{(\sigma^{-1}x + 3C \log N)\sigma^{-1} \mu} \frac{\sigma x + 3C\sigma^2 \log N}{N^{3/2}\mu^2} + 1 - e^{2(\sigma^{-1}x + 3C \log N)\sigma^{-1}\mu}. \quad (3.7)$$
The second last inequality dropped the denominator $2\pi s^3 \geq 1$ and the term $-\frac{1}{2} b(N)^2 s^{-1}$ from the exponent, and then integrated. The last inequality substituted in $z = \lambda^{-1} \log N \leq C \log N$ to bound

$$N^{-1/2} (\sigma^{-1} x + \log N) \leq b(N) \leq N^{-1/2} (\sigma^{-1} x + 3C \log N).$$

The conclusion (3.2) follows from substituting into (3.5) the bounds from above. □

4. Auxiliary facts

Before starting the proof proper, we record some simple facts. First, assumptions (2.1) and (2.2) gives these bounds:

$$0 < \sigma_\ast^2 \leq \mu_{2,N} \equiv \mathbb{E}[ (X_1^N)^2 ] = M_N^{(2)}(0) \leq C_M,$$

$$|\mu_{3,N}| \equiv |\mathbb{E}[ (X_1^N)^3 ]| = |M_N^{(3)}(0)| \leq C_M,$$

$$\mathbb{P}(X_1^N > t) \leq C_M e^{-\theta_1 t}.$$ (4.1)

**Lemma 4.1.** Let $\{Y_i\}$ be i.i.d. random variables with common marginal distribution $\nu$. Assume that, for two constants $0 < c_1, C_1 < \infty$,

$$\mathbb{E}(e^{t Y_i}) \leq C_1 \quad \text{for} \quad t \in [0, c_1].$$ (4.2)

Then

$$\mu_{\max} = \mu_{\max}(\nu, n) \equiv \mathbb{E}[\max\{0, Y_1, \ldots, Y_n\}] \leq c_1^{-1} \log(C_1 n + 1).$$

**Proof:** For $0 < t \leq c_1$,

$$e^{t \mu_{\max}} \leq \mathbb{E}(e^{t \nu_{\max} \in [0, c_1] Y_i}) \leq 1 + \mathbb{E}\left( \sum_{i=1}^{n} e^{Y_i} \right) = 1 + n \mathbb{E}(e^{Y_1}) \leq C_1 n + 1,$$

and the claim follows by taking $t = c_1$. □

Since $M_N'' > 0$ there is a unique minimizer

$$\theta_0^N = \arg \min\{M_N(\theta)\}. \quad (4.3)$$

**Lemma 4.2.** Let $N_0$ be such that $C_M |\mu_N| \leq \frac{1}{3} \sigma_\ast^2$ for $N \geq N_0$ and set $c_M = 2\sigma_\ast^{-2}$. Then for $N \geq N_0$,

$$0 \leq \theta_0^N \leq c_M |\mu_N|.$$ (4.4)

**Proof:** If $M_N'(0) = \mu_N = 0$ then the minimum is taken at $\theta_0^N = 0$.

So suppose $M_N'(0) = \mu_N < 0$. Expansion for $\theta \in (0, \theta_1)$ gives, with some $\theta' \in (0, \theta)$,

$$M_N'(\theta) = \mu_N + \mu_{2,N} \theta + \frac{1}{2} M_N^{(3)}(\theta') \theta^2 \geq \mu_N + \mu_{2,N} \theta - \frac{1}{2} C_M \theta^2.$$

Since $M'$ is strictly increasing and $c_M = 2/\sigma_\ast^2 \geq 2\mu_{2,N}^{-1}$, by the choice of $N_0$ we have for $N \geq N_0$

$$M_N'(-c_M |\mu_N|) \geq M_N' \left( -\frac{2 \mu_N}{\mu_{2,N}} \right) \geq -\mu_N - 2C_M \frac{\mu_N^2}{\mu_{2,N}^2} \geq -\mu_N (1 - \frac{3}{2} \sigma_\ast^2/\mu_{2,N}^2) > 0.$$

It follows that there exists a unique $\theta_0^N \in (0, c_M |\mu_N|)$ such that $M_N'(\theta_0^N) = 0$. □
Define a tilted measure $Q(d\omega) = f_{N,n}^{0N}(\omega)\mathbb{P}(d\omega)$ in terms of the Radon-Nikodym derivative

$$f_{N,n}^{0N}(\omega) = \frac{e_{0N}^N S_0^N}{E(e_{0N}^N S_0^N)}.$$ 

Denote the expectation under $Q$ by $\mathbb{E}^Q$. Increase $N_0$ further so that $N \geq N_0$ implies $\theta_0^N \in [-\theta_1/2, \theta_1/2]$ and $-\mu_N \leq 2$. Then for $0 \leq i \leq 3$ and $\theta \in (-\theta_1/2, \theta_1/2)$, the MGF under $Q$ satisfies

$$M_{Q,N}^{(i)}(\theta) = \mathbb{E}^Q((X_1^N)^i) = M_N^{(i+1)}(\theta_0^N) = M_N^{(i)}(\theta_0^N + \theta) \leq e^{-\mu_N\theta_0^N} C_M \leq e^{\theta_1} C_M,$$

(4.4)

where the first inequality used Jensen’s inequality and (2.1). From this we get moment bounds under $Q$: for $0 \leq i \leq 3$,

$$\mathbb{E}^Q((X_1^N)^i) = M_{Q,N}^{(i)}(0) \leq e^{\theta_1} C_M.$$ 

(4.5)

For $|\theta| \leq \theta_1$, there exists $\theta' \in (-\theta_1, \theta_1)$

$$M_N^{(2)}(\theta) = \mu_{2,N} + M_N^{(3)}(\theta') \theta$$

Increase $N_0$ further if necessary so that $\theta_0^N \leq \frac{\sigma^2}{2C_M}$ for $N \geq N_0$ and we can write

$$M_N^{(2)}(\theta_0^N) \geq \sigma^2 - C_M \theta_0^N \geq 2 \sigma^2.$$ 

Then from $\mathbb{E}^Q(X_1^N) = 0$ and the third equation in (4.4),

$$\mathbb{V}ar^Q(X_1^N) = \mathbb{E}^Q((X_1^N)^2) = M_{Q,N}^{(2)}(0) = M_N^{(2)}(\theta_0^N) \geq C_M^{-1} \sigma^2.$$ 

(4.6)

5. Proof of the main theorem

To lighten the notation we omit the label $N$ from $\mu = \mu_N$ and $\theta_0 = \theta_0^N$, and from some other notation that obviously depend on $N$. For $y > 0$ let

$$\tau_y = \inf\{m \geq 1 : |S_m^N| \geq y\}$$

denote the first hitting time of the cylinder of width $2y$. Let $\Phi_{\sigma^2}$ denote the centered Gaussian distribution with variance $\sigma^2$.

Lemma 5.1. For real $k \geq 0$ and $y \geq y_0$ we have $\mathbb{P}(\tau_y > k y^2) \leq 2 e^{-c_\tau k}$, where

$$y_0 = 1 + \sqrt{\frac{6 C_M \sigma^3}{1 - \Phi_{\sigma^2}|[-2,2]|}} \quad \text{and} \quad c_\tau = \log \left( \frac{2}{1 + \Phi_{\sigma^2}[2,\infty)} \right) \in (0, \log 2).$$

(5.1)

Proof: Let $S_m^N = S_m^N - m \mu$ be the centered walk. Consider an integer $k \geq 1$ and a real $y \geq 1$. Look at the process along time increments of size $|y|^2$:

$$\mathbb{P}(\tau_y > k y^2) \leq \mathbb{P}(\tau_y > k |y|^2) \leq \mathbb{P}(|S_m^N| \leq y \text{ for } m = 1, \ldots, k)$$

$$\leq \mathbb{P}(|S_m^N - S_{m-1}| \leq 2y \text{ for } m = 1, \ldots, k)$$

$$= \left( \mathbb{P}(S_{m-1}^N \in [-2y, 2y]) \right)^k = \left( \mathbb{P}(|y|^{-1} S^N_{m-1} \in [-2 - \mu(y), 2 - \mu(y)]) \right)^k$$

$$\leq \left( \Phi_{\sigma^2}[2,\infty - \mu(y), 2 - \mu(y)] + 3 \frac{\mu_{2,N}}{\sigma^2} |y|^{-1} \right)^k \leq \left( \Phi_{\sigma^2}[2,\infty - 2,2] + 6 C_M \sigma^3 y^{-1} \right)^k.$$
Lemma 5.2. and $y$ is, up to time $\varepsilon H$ we see that the left hand side of (5.3) is dominated by the first term on the right hand side. That is, 

$$\sup_{\sigma^2 \leq \varepsilon^2 \leq C_M} \Phi_{\sigma^2}[-2,2] + 3C_M\sigma^{-3}y^{-1} \leq \Phi_{\sigma^2}[-2,2] + 3C_M\sigma^{-3}y_0^{-1} \leq \frac{1}{2}(1 + \Phi_{\sigma^2}[-2,2]) = e^{-c_\varepsilon}. $$

We have proved $\mathbb{P}(\tau_y > ky^2) \leq e^{-c_\varepsilon k}$ for $k \in \mathbb{Z}_{\geq 0}$. Extend this to real $k \in \mathbb{R}_{\geq 0}$:

$$\mathbb{P}(\tau_y > ky^2) \leq \mathbb{P}(\tau_y > [k]y^2) \leq e^{-c_\varepsilon [k]} \leq e^{-c_\varepsilon (k-1)} = 2(1 + \Phi_{\sigma^2}[-2,2])^{-1}e^{-c_\varepsilon k} < 2e^{-c_\varepsilon k}. \quad \square$$

Let $H_N = \mu^{-2} \land N$. By (2.3)

$$D_\mu^{-2}(\log N)^6 \leq H_N \leq N. \quad (5.2)$$

Define the truncated version of $\tau_y$

$$\tilde{\tau}_y = \tau_y \land H_N.$$  

The following result shows that although the random walk $S_N^N$ has negative drift, up to times of order $H_N$ it behaves similarly to an unbiased random walk in the following sense: if $y > 0$ is not too small, but small compared to $H_N^{1/2}$, the probability that the random walk reaches level $y$ before level $-y$ is close to 1/2. Our choice of $H_N$ can be justified by decomposing the random walk into

$$S_n^N = \sum_{i=1}^n (X_i^N - \mu) + n\mu.$$ 

For $\varepsilon > 0$ small and $|\mu| \geq N^{-1/2}$ (so that $H_N = \mu^{-2}$),

$$(\varepsilon H_N)^{-1/2}S_{\varepsilon H_N}^N = (\varepsilon H_N)^{-1/2} \sum_{i=1}^{\varepsilon H_N} (X_i^N - \mu) + \varepsilon^{1/2}. \quad (5.3)$$

As

$$(\varepsilon H_N)^{-1/2} \sum_{i=1}^{\varepsilon H_N} (X_i^N - \mu) \overset{d}{\approx} N(0, \sigma),$$

we see that the left hand side of (5.3) is dominated by the first term on the right hand side. That is, up to time $\varepsilon H_N$ the random walk $S_N^N$ behaves approximately like an unbiased random walk.

**Lemma 5.2.** Let $y_0$ be as in (5.1). There exist finite constants $N_0$ and $C_0$ such that, for $N \geq N_0$ and $y_0 \leq y \leq (\log N)^{-1}H_N^{1/2}$,

$$\mathbb{P}(S_{\tilde{\tau}_y} \geq y) \geq \frac{1}{2} \left[ 1 - C_0 H_N^{-1/2}(y + (\log H_N)^2) - \frac{2}{\theta_1 y} \log(e^{\theta_1 C_M H_N} + 1) \right].$$

$C_0$ depends on $\theta_1, \sigma^2_\ast$ and $C_M$ while $N_0$ depends on $\theta_1, \sigma^2_\ast, D_\mu$ and $C_M$.

**Proof:** The constant $C_0$ comes as follows in terms of the constants previously introduced above and new constants $C_2, C_3, C_4$ introduced below in the course of the proof:

$$C_0 = C_2 + C_4 = 2(\sigma^2_\ast^{-1} + 9e^{\theta_1 \sigma^2_\ast^{-3} C_M^{5/2}}) + 2C_3 c^{-1} + 6c_M$$

$$= 2(\sigma^2_\ast^{-1} + 9e^{\theta_1 \sigma^2_\ast^{-3} C_M^{5/2}}) + 2e^{2C_M^2 c^{-1} + 4c_M c^{-1}} + 6c_M. \quad (5.4)$$
Under the measure \( Q \), \( S_n \) is a mean-zero random walk and hence a martingale. Furthermore, \( \hat{\tau}_y \) is a bounded stopping time. From this,

\[
0 = \int S_{\hat{\tau}_y} dQ = \int_{S_{\hat{\tau}_y} \geq y} S_{\hat{\tau}_y} dQ + \int_{S_{\hat{\tau}_y} \leq -y} S_{\hat{\tau}_y} dQ + \int_{S_{\hat{\tau}_y} \in (-y, y)} S_{\hat{\tau}_y} dQ.
\]

On the event \( S_{\hat{\tau}_y} \geq y \), we have \( \hat{\tau}_y = \tau_y \) and \( S_{\hat{\tau}_y-1} < y \leq S_{\hat{\tau}_y} = S_{\hat{\tau}_y-1} + X_N \leq y + X_N \) and so

\[
\int_{S_{\hat{\tau}_y} \geq y} S_{\hat{\tau}_y} dQ \leq \int_{S_{\hat{\tau}_y} \geq y} (y + X_N) dQ \leq \int_{S_{\hat{\tau}_y} \geq y} (y + 0 \vee \max_{1 \leq i \leq H_N} X_N^i) dQ \leq yQ(S_{\hat{\tau}_y} \geq y) + \mu_{\max}(Q, H_N) \leq yQ(S_{\hat{\tau}_y} \geq y) + 2\theta_1^{-1} \log(e^{\theta_1}C_M H_N + 1),
\]

where we applied Lemma 4.1 under the distribution \( Q \) with \( C_1 = e^{\theta_1}C_M, c_1 = 1/2 \theta_1 \) from (4.4). Combine the displays above to obtain

\[
Q(S_{\hat{\tau}_y} \geq y) \geq -y^{-1} \int_{S_{\hat{\tau}_y} \leq -y} S_{\hat{\tau}_y} dQ - y^{-1} \int_{S_{\hat{\tau}_y} \in (-y, y)} S_{\hat{\tau}_y} dQ - 2\theta_1^{-1} y^{-1} \log(e^{\theta_1}C_M H_N + 1)
\]

\[
\geq Q(S_{\hat{\tau}_y} \leq -y) - Q(S_{\hat{\tau}_y} \in (-y, y)) - 2\theta_1^{-1} y^{-1} \log(e^{\theta_1}C_M H_N + 1).
\]

Use

\[
Q(S_{\hat{\tau}_y} \leq -y) = 1 - Q(S_{\hat{\tau}_y} \geq y) - Q(S_{\hat{\tau}_y} \in (-y, y))
\]

to rewrite the above as

\[
Q(S_{\hat{\tau}_y} \geq y) \geq \frac{1}{2} \left[ 1 - 2Q(S_{\hat{\tau}_y} \in (-y, y)) - 2\theta_1^{-1} y^{-1} \log(e^{\theta_1}C_M H_N + 1) \right]. \tag{5.5}
\]

It remains to bound the probability on the right. \( S_{\hat{\tau}_y} \in (-y, y) \) forces \( \hat{\tau}_y = H_N \) and thereby another application of the Berry-Esseen theorem, while using (4.5), (4.6) and \( y \geq y_0 \geq 1 \), gives

\[
Q\{S_{\hat{\tau}_y} \in (-y, y)\} = Q\{H_N^{-1/2} S_{H_N} \in (-H_N^{-1/2} y, H_N^{-1/2} y)\}
\]

\[
\leq \Phi_{\sigma^*_n}(-H_N^{-1/2} y, H_N^{-1/2} y) + 3 \frac{e^{\theta_1}C_M}{2^{3/2}C_M^{-3/2}\sigma^*_n} H_N^{-\frac{1}{2}}
\]

\[
\leq 2(2\pi \sigma^*_n)^{-1/2} y H_N^{1/2} + 9 \frac{e^{\theta_1}C_M}{C_M^{-3/2}\sigma^*_n} H_N^{-\frac{1}{2}} \leq (\sigma^*_n^{-1} + 9 e^{\theta_1} \sigma^*_n^{-3} C_M^{5/2}) y H_N^{-\frac{1}{2}}
\]

\[
= \frac{1}{2} C_2 y H_N^{-\frac{3}{2}}.
\]

Rewrite (5.5) as

\[
Q(S_{\hat{\tau}_y} \geq y) \geq \frac{1}{2} \left[ 1 - C_2 y H_N^{-\frac{3}{2}} - 2y^{-1}\theta_1^{-1} \log(e^{\theta_1}C_M H_N + 1) \right]. \tag{5.6}
\]

It remains to switch from \( Q \) back to the original distribution \( \mathbb{P} \). Recall the Radon-Nikodym derivative \( f_0^n = M(\theta)^{-n} e^{\theta S_n} \). Introduce a temporary quantity \( G_0 > 1 \) to be chosen precisely below. Decompose according to the value of \( \hat{\tau}_y \) and use Cauchy-Schwarz:

\[
Q(S_{\hat{\tau}_y} \geq y) = \mathbb{E}[f_0^n (1_{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y \leq G_0} + 1_{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0})]
\]

\[
\leq \mathbb{E}[f_0^n 1_{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y \leq G_0}] + \left( \mathbb{E}[f_0^n]^2 \right)^{\frac{1}{2}} \left( \mathbb{P}\{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0\} \right)^{\frac{1}{2}}. \tag{5.7}
\]
Let us first bound the second term on line (5.7). Note that $f_n^\theta$ is a $\mathbb{P}$-martingale and $\hat{\tau}_y$ is a stopping time bounded by $H_N$. Hence $(f_n^\theta)^2$ is a submartingale and we have

$$(\mathbb{E}[(f_{\hat{\tau}_y}^\theta)^2])^{\frac{1}{2}} \left( \mathbb{P}\{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0\} \right)^{\frac{1}{2}} \leq \left( \mathbb{E}\{(f_{H_N}^\theta)^2\} \right)^{\frac{1}{2}} \left( \mathbb{P}\{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0\} \right)^{\frac{1}{2}} = \left( \frac{M(2\theta_0)}{M(\theta_0)^2} \right)^{\frac{H_N}{2}} \left( \mathbb{P}\{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0\} \right)^{\frac{1}{2}}.$$ 

To bound the $M$-factor on the right, expand $M$ and use (2.1), (4.1) and $\mu < 0$. In the numerator, for some $\eta \in (0,2\theta_0)$,

$$M(2\theta_0) = 1 + \mu 2\theta_0 + 2\mu_2 \theta_0^2 + \frac{6}{5} M^{(3)}(\eta) \theta_0^3 \leq 1 + 2\mu_2 \theta_0^2 + \frac{4}{5} C_M \theta_0^3$$

and similarly in the denominator:

$$\left[ M(2\theta_0) M(\theta_0)^{-2} \right]^{\frac{H_N}{2}} \leq \left( 1 + 2\mu_2 \theta_0^2 + \frac{4}{5} C_M \theta_0^3 \right)^{\frac{H_N}{2}} \left( 1 + \mu \theta_0 + \frac{1}{2} \mu_2 \theta_0^2 - \frac{1}{6} C_M \theta_0^3 \right)^{-H_N} \leq \left( 1 + 2 C_M \mu^2 + \frac{4}{3} C_M^3 |\mu|^3 \right)^{\frac{1}{2} \mu^{-2}} \left( 1 - c_M \mu^2 - C_M c_M^3 |\mu|^3 \right)^{-\mu^{-2}}$$

where the second line also used $(1-a)^{-1} \leq 1 + 2a$ for $a \in [0,\frac{1}{2}]$. Put (5.8) back up, set $G_0 = y H_N^{1/2}$, and apply Lemma 5.1 (for which we use the assumption $y \geq y_0$):

$$\left( \mathbb{E}[(f_{\hat{\tau}_y}^\theta)^2] \right)^{\frac{1}{2}} \left( \mathbb{P}\{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y > G_0\} \right)^{\frac{1}{2}} \leq C_3 \left( \mathbb{P}\{\hat{\tau}_y > G_0\} \right)^{\frac{1}{2}} \leq 2 C_3 e^{-c_H H_N^{1/2} y^{-1}}.$$ 

Next we bound the first term on line (5.7). Use $M(\theta_0) \leq 1$. Let $\mathcal{M}_n = \max_{1 \leq i \leq n} X_i^N$.

\[
\mathbb{E}\left[ f_{\hat{\tau}_y}^\theta \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y \leq G_0} \right] = \mathbb{E}\left[ \frac{e^{\theta_0 S_{\hat{\tau}_y}}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \hat{\tau}_y \leq G_0} \right] \leq \mathbb{E}\left[ \frac{e^{\theta_0 S_{\hat{\tau}_y}}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \mathcal{M}_N \leq (\log H_N)^2} \right] \mathbb{E}\left[ \frac{e^{\theta_0 S_{\hat{\tau}_y}}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \mathcal{M}_N > (\log H_N)^2} \right] \leq \mathbb{E}\left[ \frac{e^{\theta_0 (y + \mathcal{M}_N)}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \mathcal{M}_N \leq (\log H_N)^2} \right] \mathbb{E}\left[ \frac{e^{\theta_0 S_{\hat{\tau}_y}}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \mathcal{M}_N > (\log H_N)^2} \right].
\]

Let us first bound the second term. Using Cauchy-Schwarz, the bound

$$\mathbb{P}(\mathcal{M}_N > t) \leq H_N C_M e^{-\theta_1 t},$$

the bound (5.8), and the tail bound in (4.1), it follows that

\[
\mathbb{E}\left[ \frac{e^{\theta_0 S_{\hat{\tau}_y}}}{M(\theta_0)^2} \mathbb{I}_{S_{\hat{\tau}_y} \geq y, \mathcal{M}_N > (\log H_N)^2} \right] \leq \left( \mathbb{E}\left[ (f_{H_N}^\theta)^2 \right] \right)^{\frac{1}{2}} \left( \mathbb{P}\{\mathcal{M}_N > (\log H_N)^2\} \right)^{\frac{1}{2}} \leq C_3 H_N^{1/2} C_M^{-1/2} e^{-\frac{1}{2} \theta_1 (\log H_N)^2}.
\]
The first term on the last line of (5.10) is bounded as follows, with $G_0 = yH_N^{1/2}$.

$$
\mathbb{E}\left[ \frac{e^{\theta_0(y+\mathcal{M}H_N)}}{M(\theta_0)^{G_0}} 1_{S_{T_y} \geq y, \mathcal{M}H_N \leq (\log H_N)^2} \right] \leq \mathbb{E}\left[ \frac{e^{\theta_0(y+(\log H_N)^2)}}{M(\theta_0)^{G_0}} 1_{S_{T_y} \geq y} \right]
$$

$$
\leq \mathbb{P}(S_{T_y} \geq y) e^{c_M H_N^{-1/2} y + (\log H_N)^2} M(\theta_0)^{-y H_N^{1/2}} \leq \mathbb{P}(S_{T_y} \geq y) e^{c_M H_N^{-1/2} y + (\log H_N)^2} e^{c_M y H_N^{-1/2}}
$$

$$
= \mathbb{P}(S_{T_y} \geq y) e^{c_M H_N^{-1/2} (2y + (\log H_N)^2)}
$$

$$
\leq \mathbb{P}(S_{T_y} \geq y) \left[ 1 + 2c_M H_N^{-1/2} [2y + (\log H_N)^2] \right] \leq \mathbb{P}(S_{T_y} \geq y) + 2c_M H_N^{-1/2} [2y + (\log H_N)^2].
$$

We used above Jensen’s inequality in the form $M(\theta_0)^{-y H_N^{1/2}} \leq e^{-\theta_0 \mu H_N^{1/2}}$, the definition of $H_N$ in the form $|\mu| H_N^{1/2} \leq 1$, and then $\theta_0 \leq c_M |\mu| \leq c_M H_N^{-1/2}$. Furthermore, by (5.2) and our assumption $y \leq (\log N)^{-1} H_N^{1/2}$ we have

$$
c_M H_N^{-1/2} [2y + (\log H_N)^2] \leq c_M (2 + D_\mu)(\log N)^{-1} \leq 2
$$

where we choose $N_0$ large enough so that the last inequality holds for $N \geq N_0$. Then we applied the inequality $e^x \leq 1 + 2x$ for $x \in [0, \log 2]$. Going back to (5.10), for $N \geq N_0$,

$$
\mathbb{E}\left[ f_{\bar{T}_y} 1_{S_{T_y} \geq y, \bar{T}_y \in G_0} \right] \leq \mathbb{P}(S_{T_y} \geq y) + 2c_M H_N^{-1/2} [2y + (\log H_N)^2] + C_3 H_N^{1/2} C_M^{1/2} e^{-\frac{1}{2} \theta_1 (\log H_N)^2}
$$

$$
\leq \mathbb{P}(S_{T_y} \geq y) + 3c_M H_N^{-1/2} [2y + (\log H_N)^2].
$$

The second inequality is guaranteed for example by choosing $N_0$ large enough so that $N \geq N_0$ implies

$$
D_\mu^{-2}(\log N)^6 \geq e^{\theta_1^{-1}} \quad \text{and} \quad c_M C_3^{-1} C_M^{-1/2} (\log [D_\mu^{-2}(\log N)^6])^2 \geq e^{\frac{3}{2} \theta_1^{-1}}.
$$

This works due to the lower bound (5.2) on $H_N$ and because the function $f(x) = xe^{-\frac{1}{2} \theta_1 (\log x)^2}$ achieves its maximum $e^{\frac{3}{2} \theta_1^{-1}}$ at $x = e^{\theta_1^{-1}}$ after which it decreases.

Combine the above with (5.9) on line (5.7) to get this upper bound:

$$
Q(S_{T_y} \geq y) \leq \mathbb{P}(S_{T_y} \geq y) + 2C_3 e^{-c_r H_N^{1/2} y^{-1}} + 3c_M H_N^{-1/2} [2y + (\log H_N)^2]
$$

$$
\leq \mathbb{P}(S_{T_y} \geq y) + 2C_3 e^{-c_r H_N^{1/2} y^{-1}} + 3c_M H_N^{-1/2} [2y + (\log H_N)^2]
$$

$$
\leq \mathbb{P}(S_{T_y} \geq y) + C_4 H_N^{-1/2} [y + (\log H_N)^2]
$$

where $C_4 = 2C_3 e^{-c_r^{-1}} + 6c_M$. The second inequality above came from $xe^{-x} \leq e^{-1}$ for $x \geq 0$. Put (5.11) and (5.6) together to obtain the claim of the lemma.

By adjusting a constant we can replace $\bar{T}_y$ with $T_y$ in the previous estimate.

**Corollary 5.3.** Under the assumptions of Lemma 5.2, with $C_{10} = C_0 + 2c_r^{-1}$,

$$
\mathbb{P}(S_{T_y} \geq y) \geq \frac{1}{2} \left[ 1 - C_10 H_N^{-1} (y + (\log H_N)^2) - \frac{2}{\theta_1 y} \log (e^{\theta_1} C_M H_N + 1) \right] \quad (5.12)
$$

**Proof:** The assumption $y_0 \leq y \leq H_N^{1/2}$ implies that Lemma 5.1 applies to give

$$
\mathbb{P}(T_y > H_N) \leq e^{-c_r H_N y^{-2}} \leq e^{-c_r H_N^{1/2} y^{-1}} \leq \theta_1 H_N^{-1/2} y.
$$

The claim then comes from Lemma 5.2 and $\mathbb{P}(S_{T_y} \geq y) \geq \mathbb{P}(S_{T_y} \geq y) - \mathbb{P}(T_y > H_N)$.

For $w > 0$ truncate:

$$
\hat{X}_i^{N,w} = X_i^N 1_{\{x_i^N \geq -w\}} - w 1_{\{x_i^N < -w\}} \quad \text{and} \quad \hat{S}_n^{N,w} = \sum_{i=1}^n \hat{X}_i^{N,w}.
$$
Define
\[ t_y = \inf\{ m \geq 1 : |\hat{S}_m^{N,w}| \geq y \}. \]

We transfer bound (5.12) to the truncated walk \( \hat{S} \). The reason is that the proof of the forthcoming Lemma 5.5 is easier for the truncated RW.

Corollary 5.4. Under the assumptions of Lemma 5.2, with \( C_{11} = C_0 + 4c_\tau^{-1} \),
\[ \mathbb{P}(\hat{S}_{t_y}^{N,w} \geq y) \geq \frac{1}{2} \left[ 1 - C_{11} H_N^{-\frac{1}{2}} (y + (\log H_N)^2) - \frac{2}{\theta_1 y} \log(e^{\theta_1} C_M H_N + 1) - H_N C_M e^{-\theta_1 w} \right]. \]

Proof: Note that
\[ \mathbb{P}(\hat{S}_m^{N,w} \neq S_m^N \text{ for some } 1 \leq m \leq H_N) = \mathbb{P}(\hat{X}_i^{N,w} \neq X_i^N \text{ for some } 1 \leq i \leq H_N) \]
\[ = \mathbb{P}\left( \inf_{1 \leq i \leq H_N} X_i^N < -w \right) \leq H_N C_M e^{-\theta_1 w}. \]

Moreover,
\[ \mathbb{P}(\hat{S}_{t_y}^{N,w} \geq y) \geq \mathbb{P}(S_{t_y}^{N} \geq y, \tau_y \leq H_N, \hat{S}_{t_y}^{N,w} = S_{t_y}^N \text{ for all } 1 \leq m \leq H_N) \]
\[ \geq \mathbb{P}(S_{t_y}^{N} \geq y) - \mathbb{P}(\tau_y > H_N) - \mathbb{P}(\hat{S}_{t_y}^{N,w} \neq S_{t_y}^N \text{ for some } 1 \leq m \leq H_N) \]
\[ \geq \mathbb{P}(S_{t_y}^{N} \geq y) - c_\tau^{-1} H_N^{-1/2} y - H_N C_M e^{-\theta_1 w} , \]
where we used (5.14) and (5.13). Combine the above with (5.12) to obtain the result. \( \square \)

We turn to the main argument of the proof of Theorem 2.2, that is, to show that the probability of the random walk \( \hat{S}_m \) to hit the level \( x \) before hitting the level \( -\varepsilon H_1^{1/2} \) is close to \( x/|\mu| \). This gives rise to the error term in (2.4). We sketch the reasoning.

Let us try to hit the level \( x > 0 \) starting from the origin. By Corollary 5.4 there is a probability \( \approx 1/2 \) to hit \( x \) before hitting \( -x \). Suppose we failed and hit \( -x \) first. We have another chance to hit \( x \) by going \( 2x \) upward from the level \( -x \). By Corollary 5.4 the probability of going \( 2x \) up to the level \( x \) before going \( 2x \) down to the level \( -4x \) is \( \approx 1/2 \). We continue this way until we either hit the level \( x \) or the level \( -\varepsilon H_1^{1/2} \). How many trials to hit \( x \) do we have before we hit \( -\varepsilon H_1^{1/2} \)? Approximately \( K = \log_2(x^{-1}\varepsilon H_1^{1/2}) \). The trials are independent and so the probability of hitting the level \( -\varepsilon H_1^{1/2} \) before hitting the level \( x \) is \( \approx 2^{-K} = Cx|x| \), which is what we seek.

We introduce the notation to make the sketch precise. See Figure 5.1 for an illustration.

Define \( K = \lfloor \log_2(x^{-1}(\log N)^{-1} H_1^{1/2}) \rfloor \). For \( i \geq 0 \) set \( L_i = 2^{i+2} - 3 \). Inductively these satisfy \( L_0 = 1 \) and \( L_i = 2L_{i-1} + 3 \). Furthermore,
\[ xL_K \leq (\log N)^{-1} H_1^{1/2}. \]

Define the stopping times
\[ T_0 = \inf\{ n : |\hat{S}_n^{N,x}| \geq xL_0 \} \]
and
\[ T_i = \inf\{ n \geq T_{i-1} : \hat{S}_n^{N,x} \leq -xL_i \text{ or } \hat{S}_n^{N,x} \geq x \}. \]

Note that \( T_i = T_{i-1} \) is possible.

Lemma 5.5. There exist finite constants \( C_{12} \) and \( N_0 \) such that for \( N \geq N_0 \) and \( x \geq (\log N)^2 \),
\[ \mathbb{P}\left( \max_{1 \leq m \leq T_K} \hat{S}_m^{N,x} < x \right) \leq C_{12} x (\log N) H_1^{-1/2}, \]
where
\[ C_{12} = 4 \exp\{ 4(C_0 + 4c_\tau^{-1})(1 + D_\mu) + 8\theta_1^{-1} (1 + \log(e^{\theta_1} C_M + 1)) + 4C_M \} \]
and \( C_0 \) in the expression above is from (5.4).
Proof: Since $C_0 \geq 2$, we have $C_{12} \geq 4e^8 \geq 2^{10}$. Then we can assume that $x \leq 2^{-10} (\log N)^{-1} H_N^{1/2}$, for otherwise the bound on the probability is $> 1$. This guarantees that $K \geq 8$. It also implies that unless $|\mu| \leq 2^{-10} (\log N)^{-3}$, the result is trivial.

Since $\hat{N}_{i,x}^{N,x} \geq -x$,

$$\{\hat{S}_{T_i}^{N,x} \leq -xL_i\} = \{-x(L_i + 1) < \hat{S}_{T_i}^{N,x} \leq -xL_i\}$$

and

$$\{\hat{S}_{T_i}^{N,x} \leq -xL_i, \hat{S}_{T_{i+1}}^{N,x} \leq -xL_{i+1}\} \subseteq \{T_i < T_{i+1}\}.$$ 

Note that

$$E \equiv \left\{ \max_{1 \leq m \leq T_K} \hat{S}_m^{N,x} < x \right\} \subseteq \bigcap_{1 \leq i \leq K} \{\hat{S}_{T_i}^{N,x} \leq -xL_i\}. \quad (5.17)$$

Due to (5.16)

$$\mathbb{P}\left(\hat{S}_{T_0}^{N,x} \leq -xL_0, \ldots, \hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}, \hat{S}_{T_i}^{N,x} \leq -xL_i\right)$$

$$= \mathbb{P}\left(\hat{S}_{T_i}^{N,x} \leq -xL_i | \hat{S}_{T_0}^{N,x} \leq -xL_0, \ldots, \hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}\right) \cdot \mathbb{P}\left(\hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}\right)$$

$$\leq \mathbb{P}\left(\hat{S}_{T_i}^{N,x} \leq -xL_i | \hat{S}_{T_0}^{N,x} \leq -xL_0, \ldots, \hat{S}_{T_{i-1}}^{N,x} = -x(L_{i-1} + 1)\right) \cdot \mathbb{P}\left(\hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}\right)$$

$$= \mathbb{P}\left(\hat{S}_{T_i}^{N,x} \leq -xL_i, \hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}\right) \mathbb{P}\left(\hat{S}_{T_{i-1}}^{N,x} \leq -xL_{i-1}\right). \quad (5.18)$$

The last equality used the definition of the stopping time $t_y$, the definition of $L_i$, and the Markov property. For $1 \leq i \leq K$ define the events

$$A_i^N = \{\hat{S}_{t_{y(L_i-1) + 2}}^{N,x} \leq -x(L_{i-1} + 2)\}.$$ 

Applying (5.18) to (5.17) repeatedly,

$$\mathbb{P}(E) \leq \mathbb{P}\left(\bigcap_{1 \leq i \leq K} \{\hat{S}_{T_i}^{N,x} \leq -xL_i\}\right) \leq \prod_{1 \leq i \leq K} \mathbb{P}(A_i^N).$$

Let $x \geq (\log N)^2$. Recall that by (5.2), $(\log H_N)^2 H_N^{-1/2} \leq D(\log N)^{-1}$ and $H_N \leq N$. Apply Corollary 5.4 with $w = x$ and $y_i = x(L_{i-1} + 2) \in [x, (\log N)^{-1} H_N^{1/2}]$ for $i = 1, \ldots, K$ and $N \geq N_0$ to get this estimate:

$$\mathbb{P}(A_i^N) \leq \frac{1}{2} \left[1 + C_{11} H_N^{-\frac{1}{2}} (y_i + (\log H_N)^2) + \frac{2}{\theta_1 y_i} \log(e^{\theta_1} C_M H_N + 1) + H_N C_M e^{-\theta_1 x}\right]$$

$$\leq \frac{1}{2} \left[1 + C_{11} (1 + D(\log N)^{-1}) + \frac{2\log(e^{\theta_1} C_M N + 1)}{\theta_1 (\log N)^2} + C_M N^{1-\theta_1 (\log N)^2}\right]$$

$$\leq \frac{1}{2} \left[1 + C_A (\log N)^{-1}\right],$$

where we set

$$C_A = C_{11} (1 + D) + 2\theta_1^{-1} (1 + \log(e^{\theta_1} C_M + 1)) + C_M$$

and if necessary we increase $N_0$ further so that $N^{1-\theta_1 (\log N)^2} \leq (\log N)^{-1}$ for $N \geq N_0$. 

By the time the random walk \( \hat{S}^N \) exits the cylinder of radius \((\log N)^{-1} H_N^{1/2}\) it has had about \(K\) independent opportunities to hit the level \(x\), each with probability close to \(\frac{1}{2}\).

Continue with the above estimate,

\[
\Pr(E) \leq \prod_{i=1}^{K} \Pr(A_i^N) \leq \left( \frac{1}{2} [1 + C_A(\log N)^{-1}] \right)^K
\]

\[
= \left( \frac{1}{2} [1 + C_A(\log N)^{-1}] \right)^{\log_2(x^{-1}(\log N)^{-(1/2)}H_N^{1/2})-2}
\]

\[
\leq 4x(\log N)H_N^{-1/2}[1 + C_A(\log N)^{-1}]^{\log_2 N}
\]

\[
\leq 4e^{4C_A}x(\log N)H_N^{-1/2} = 4e^{4C_A}(\log N)(|\mu| \lor N^{-1/2}),
\]

where we used \(\log_2 N = \frac{\log N}{\log 2} \leq 4\log N\).

We are ready to prove Theorem 2.2. By Lemma 5.5, by the time \(\hat{S}\) hits the level \((\log N)^{-1} H_N^{1/2}\), with high probability it has hit level \(x\) as well. It remains to verify the two points below.

(i) \(\hat{S}\) is close to \(S\) on the time interval \([1, N]\). This follows from a union bound and the exponential tail of \(X_i^N\).

(ii) With high probability by time \(N\) we hit the boundary of the cylinder of width \((\log N)^{-1} H_N^{1/2}\). This follows from Lemma 5.1.

**Proof of Theorem 2.2:** Consider \(x \geq (\log N)^2\). Observe that

\[
\left\{ \max_{1 \leq m \leq N} |\hat{S}_m^{N,x}| \geq (\log N)^{-1} H_N^{1/2}, \max_{1 \leq m \leq T_K} \hat{S}_m^{N,x} > x \right\} \subseteq \left\{ \max_{1 \leq m \leq N} \hat{S}_m^{N,x} > x \right\}.
\]

Indeed, on the event \(\hat{S}_m^{N,x} \geq (\log N)^{-1} H_N^{1/2} \geq xT_K\) we have \(T_K \leq N\).

Next,

\[
\Pr(\hat{S}_m^{N,x} \neq S_m^N \text{ for some } 1 \leq m \leq N) = \Pr(X_i^N \leq -x \text{ for some } 1 \leq i \leq N)
\]

\[
\leq C_M Ne^{-\theta_1 x} \leq C_M N e^{-\theta_1 (\log N)^2}.
\]
By Lemma 5.1, (5.19) and Lemma 5.5,
\[
\begin{align*}
\mathbb{P}( \max_{1 \leq m \leq N} \hat{S}^{N,x}_m \geq x ) & \geq 1 - \mathbb{P}( \max_{1 \leq m \leq N} \hat{S}^{N,x}_m < (\log N)^{-1} H_N^{1/2} ) - \mathbb{P}( \max_{1 \leq m \leq T_K} \hat{S}^{N,x}_m \leq x ) \\
& \geq 1 - \left[ \mathbb{P}( \max_{1 \leq m \leq N} |S^N_m| < (\log N)^{-1} H_N^{1/2} ) + \mathbb{P}( \hat{S}^{N,x}_m \neq S^N_m \text{ for some } 1 \leq m \leq N ) \right] \\
- \mathbb{P}( \max_{1 \leq m \leq T_K} \hat{S}^{N,x}_m \leq x ) \\
& = 1 - \left[ \mathbb{P}( \tau_{(\log N)^{-1} H_N^{1/2}} > N ) + \mathbb{P}( \hat{S}^{N,x}_m \neq S^N_m \text{ for some } 1 \leq m \leq N ) \right] \\
- \mathbb{P}( \max_{1 \leq m \leq T_K} \hat{S}^{N,x}_m < x ) \\
& \geq 1 - \left[ 2e^{-c_0(\log N)^2 H_N^{-1}} + C_M N e^{-\theta_1 (\log N)^2} \right] - C_1 x (\log N) H_N^{-1/2} \\
& \geq 1 - 2e^{-c_0(\log N)^2} - C_M N e^{-\theta_1 (\log N)^2} - C_1 x (\log N) H_N^{-1/2} \\
& \geq 1 - (C_1 + 2) x (\log N) H_N^{-1/2}.
\end{align*}
\]

To get the inequalities above for \( N \geq N_0 \) we increase \( N_0 \) if necessary so that \( N \geq N_0 \) guarantees \((\log N)^{-1} H_N^{1/2} \geq y_0\) to apply Lemma 5.1, and furthermore so that \( 2e^{-c_0(\log N)^2} \vee C_M N e^{-\theta_1 (\log N)^2} \leq (\log N)^3 N^{-1/2} \) to get the last inequality.

Now the final inequality:
\[
\begin{align*}
\mathbb{P}( \max_{1 \leq m \leq N} S^N_m \geq x ) & \geq \mathbb{P}( \max_{1 \leq m \leq N} \hat{S}^{N,x}_m \geq x ) - \mathbb{P}( \hat{S}^{N,x}_m \neq S^N_m \text{ for some } 1 \leq m \leq N ) \\
& \geq 1 - (C_1 + 2) x (\log N) H_N^{-1/2} - C_M N e^{-\theta_1 (\log N)^2} \\
& \geq 1 - (C_1 + 3) x (\log N)(\mu \vee N^{-1/2}).
\end{align*}
\]

Theorem 2.2 has been proved. \( \square \)

References


