# Large deviations and slowdown asymptotics for one-dimensional excited random walks[*]

Jonathon Peterson[†]

## Abstract

We study the large deviations of excited random walks on $\mathbb{Z}$. We prove a large deviation principle for both the hitting times and the position of the random walk and give a qualitative description of the respective rate functions. When the excited random walk is transient with positive speed $v_0$, then the large deviation rate function for the position of the excited random walk is zero on the interval $[0, v_0]$ and so probabilities such as $P(X_n < nv)$ for $v \in (0, v_0)$ decay subexponentially. We show that rate of decay for such slowdown probabilities is polynomial of the order $n^{1-\delta/2}$, where $\delta > 2$ is the expected total drift per site of the cookie environment.

## 1 Introduction

In this paper we study the large deviations for one-dimensional excited random walks. Excited random walks are a model for a self-interacting random walk, where the transition probabilities depend on the number of prior visits of the random walk to the current site. The most general model for excited random walks on $\mathbb{Z}$ is the following. Let $\Omega = [0, 1]^{\mathbb{Z} \times \mathbb{N}}$, and for any element $\omega = \{\omega_i(j)\}_{i \in \mathbb{Z}, j \geq 1} \in \Omega$ we can define an excited random walk $X_n$ by letting $\omega_i(j)$ be the probability that the random walk moves to the right upon its $j$-th visit to the site $i \in \mathbb{Z}$. More formally, we will let $P_\omega(X_0 = 0)$ and

$$P_\omega(X_{n+1} = X_n + 1 | \mathcal{F}_n) = 1 - P_\omega(X_{n+1} = X_n - 1 | \mathcal{F}_n) = \omega_x \left( \#\{k \leq n : X_k = x\} \right),$$

where $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. Note that the excited random walk $X_n$ is not a Markov chain since the transition probabilities depend on the entire past of the random walk and not just the current location.

Excited random walks are also sometimes called *cookie random walks*, since one imagines a stack of "cookies" at every site which each induce a specific bias to the

---

[†]Purdue University, USA. E-mail: peterson@math.purdue.edu

walker. When the walker visits the site $x$ for the $i$-th time, he eats the $i$-th cookie which causes his next step to be as a simple random walk with parameter $\omega_x(i)$. For this reason we will also refer to $\omega = \{\omega_i(j)\}_{i \in \mathbb{Z}, \, j \geq 1}$ as a *cookie environment*.

We can also assume that the cookie environment $\omega$ is first chosen randomly. That is, let $\mathbf{P}$ be a probability distribution on the space of cookie environments $\Omega$, and define a new measure on the space of random walk paths $\mathbb{Z}^{\mathbb{Z}_+}$ by averaging over all cookie environments. That is, let

$$P(\cdot) = \int_\Omega P_\omega(\cdot) \, \mathbf{P}(d\omega).$$

For a fixed cookie environment $\omega$, the law $P_\omega$ is referred to as the *quenched* law of the excited random walk, and $P$ is called the *averaged* law of the excited random walk.

Most of the results for excited random walks make the assumption that there are only finitely many cookies per site. That is, there exists an $M$ such that $\omega_i(j) = 1/2$ for any $i \in \mathbb{Z}$ and $j > M$ so that after $M$ visits to any site the transitions are like a simple symmetric random walk.

**Assumption 1.** *There exists an integer $M < \infty$ such that there are almost surely only $M$ cookies per site. That is, $\mathbf{P}(\Omega_M) = 1$, where*

$$\Omega_M = \Omega \cap \{\omega : \omega_i(j) = 1/2, \, \forall i \in \mathbb{Z}, \, \forall j > M\}.$$

We will also make the common assumption that the cookie environment is i.i.d. in the following sense.

**Assumption 2.** *The distribution $\mathbf{P}$ is such that the sequence of cookie environments at each site $\{\omega_i(\cdot)\}_{i \in \mathbb{Z}}$ is i.i.d.*

Finally, we will make the following non-degeneracy assumption on cookie environments.

**Assumption 3.** *With $M$ as in Assumption 1,*

$$E\left[\prod_{j=1}^M \omega_0(j)\right] > 0 \quad \text{and} \quad E\left[\prod_{j=1}^M (1 - \omega_0(j))\right] > 0.$$

Excited random walks were first studied by Benjamini and Wilson in [3], where they considered the case of deterministic cookie environments with one cookie per site (that is $M = 1$). The focus of Benjamini and Wilson was mainly on the $\mathbb{Z}^d$ case, but in the special case of $d = 1$ they showed that excited random walks with one cookie per site are always recurrent. The model was further generalized by Zerner in [21] to allow for multiple cookies per site and for randomness in the cookie environment, but with the restriction that all cookies induced a non-negative drift (that is $\omega_i(j) \geq 1/2$). Recently the model of excited random walks was further generalized by Zerner and Kosygina to allow for cookies with both positive and negative drifts [14].

The recurrence/transience and limiting speed for one-dimensional excited random walks have been studied in depth under the above assumptions. A critical parameter for describing the behavior of the excited random walk is the expected total drift per site

$$\delta = \mathbf{E}\left[\sum_{i \geq 1} (2\omega_0(j) - 1)\right] = \mathbf{E}\left[\sum_{i=1}^M (2\omega_0(j) - 1)\right]. \tag{1.1}$$

Zerner showed in [21] that excited random walks with all cookies $\omega_i(j) \geq 1/2$ are transient to $+\infty$ if and only if $\delta > 1$. Additionally, Zerner showed that the limiting speed

$v_0 = \lim_{n\to\infty} X_n/n$ exists, $P$-a.s., but was not able to determine when the speed is non-zero. Basdevant and Singh solved this problem in [1] where they showed that $v_0 > 0$ if and only if $\delta > 2$. These results for recurrence/transience and the limiting speed were given only for cookies with non-negative drift but were recently generalized by Kosygina and Zerner [14] to the general model we described above that allows for cookies with both positive and negative drifts. In summary, under Assumptions 1 – 3, the following results are known.

- $X_n$ is recurrent if and only if $\delta \in [-1, 1]$. Moreover,

$$\lim_{n\to\infty} X_n = \begin{cases} -\infty & \text{if } \delta < -1 \\ +\infty & \text{if } \delta > 1, \end{cases} \qquad P\text{-a.s.}$$

- There exists a constant $v_0$ such that $\lim_{n\to\infty} X_n/n = v_0$, $P$-a.s. Moreover, $v_0 = 0$ if and only if $\delta \in [-2, 2]$.

Limiting distributions for excited random walks are also known with the type of rescaling and limiting distribution depending only on the parameter $\delta$ given in (1.1). The interested reader is referred to the papers [2, 10, 11, 13, 14] for more information on limiting distributions.

## 1.1 Main Results

In this paper, we are primarily concerned with the large deviations of excited random walks. In a similar manner to the approach used for large deviations of random walks in random environments, we deduce a large deviation principle for $X_n/n$ from a large deviation principle for $T_n/n$, where

$$T_n = \inf\{k \geq 0 : X_k = n\}, \quad n \in \mathbb{Z}$$

are the hitting times of the excited random walk. However, we do not prove a large deviation principle for the hitting times directly. Instead, we use an associated branching process with migration $V_i$ that has been used previously in some of the above mentioned papers on the speed and limiting distributions for excited random walks [1, 13, 14]. We prove a large deviation principle for $n^{-1} \sum_{i=1}^{n} V_i$ and use this to deduce a large deviation principle for $T_n/n$ which in turn implies the following large deviation principle for $X_n/n$.

**Theorem 1.1.** *The empirical speed of the excited random walk $X_n/n$ satisfies a large deviation principle with rate function $I_X$ defined in (5.1). That is, for any open set $G \subset [-1, 1]$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log P(X_n/n \in G) \geq - \inf_{x\in G} I_X(x), \tag{1.2}$$

*and for any closed set $F \subset [-1, 1]$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(X_n/n \in F) \leq - \inf_{x\in F} I_X(x).$$

**Remark 1.2.** *After the initial draft of this paper was completed, it was noted that a general large deviation principle for certain non-Markovian random walks due to Rassoul-Agha [19] can be used to prove Theorem 1.1 in certain cases. Thus, it is necessary to point out some of the differences with the current paper.*

- *In [19] the random walks are assumed to be uniformly elliptic, which in the context of this paper would require $\omega_i(j) \in [c, 1-c]$ for all $i \in \mathbb{Z}$, $j \geq 1$ and some $c > 0$. In contrast, we only assume the weaker condition in Assumption 3.*

- *The results of [19] only apply directly to excited random walks with deterministic cookie environments. If the cookie environments are allowed to be random and satisfying Assumption 2, then a technical difficulty arises in satisfying one of the conditions for the large deviation principle in [19]. Specifically, the transition probabilities $q(\mathrm{w}, z)$ for the shifted paths as defined in [19] do not appear to be continuous in $\mathrm{w}$ for the required topology. We suspect, however, that the techniques of [19] could be adapted to apply to this case as well.*
- *The formulation of the large deviation rate function in [19] is difficult to work with and the only stated properties of the rate function are convexity and a description of the zero set. In contrast, our method gives a more detailed description of the rate function (see Lemma 5.1 and Figure 3).*
- *The method in [19] does not also give a large deviation principle for the hitting times of the random walk, though one could use an argument similar to that in Section 5 below to deduce a large deviation principle for the hitting times from the large deviation principle for the location of the random walk.*

As mentioned in the above remark, the formulation of the rate function $I_X$ given in the proof of Theorem 1.1 allows us to give a good qualitative description of the rate function (see Lemma 5.1). One particularly interesting property is that if $\delta > 2$ (so that the limiting speed $v_0 > 0$) then $I_X(x) = 0$ when $x \in [0, v_0]$. Thus, probabilities of the form $P(X_n < nx)$ decay subexponentially if $x \in (0, v_0)$. In fact, as the following example shows, one can see quite easily that such slowdown probabilities must have a subexponential rate of decay.

**Example 1.3.** *We exhibit a naive strategy for obtaining a slowdown of the excited random walk. Consider the event where the excited random walk first follows a deterministic path that visits every site in $[0, n^{1/3})$ $M$ times (so that no cookies remain in the interval) and then the random walk stays in the interval $[0, n^{1/3})$ for $n$ steps. The probabilistic cost of forcing the random walk to follow the deterministic path at the beginning is $e^{-c'Mn^{1/3}}$ for some $c' > 0$. Then, since there are no cookies left in the interval, the probability of then staying in $[0, n^{1/3})$ for $n$ steps before exiting to the right is a small deviation computation for a simple symmetric random walk. The probability of this event can be bounded below by $Ce^{-c''n^{1/3}}$ for some $C, c'' > 0$ (see Theorem 3 in [16]). Thus, the total probability of the above event for the excited random walk is at least $Ce^{-cn^{1/3}}$.*

The example above shows that $P(X_n < xn)$ decays slower than a stretched exponential. However, this strategy turns out to be far from the optimal way for obtaining such a slowdown. The second main result of this paper is that the true rate of decay for slowdowns is instead polynomial of the order $n^{1-\delta/2}$.

**Theorem 1.4.** *If $\delta > 2$, then*

$$\lim_{n\to\infty} \frac{\log P(X_n < nx)}{\log n} = 1 - \frac{\delta}{2}, \quad \forall x \in (0, v_0), \tag{1.3}$$

*and*

$$\lim_{n\to\infty} \frac{\log P(T_n > nt)}{\log n} = 1 - \frac{\delta}{2}, \quad \forall t > 1/v_0. \tag{1.4}$$

## 1.2 Comparison with RWRE

Many of the prior results for one-dimensional excited random walks are very similar to the corresponding statements for random walks in random environments (RWRE). For instance, both models can exhibit transience with sublinear speed and they have the

same types limiting distributions for the hitting times and the location of the random walk [14, 20, 12]. Thus, it is interesting to compare the results of this paper with what is known for one-dimensional RWRE.

Large deviations for one-dimensional RWRE (including a qualitative description of the rate functions) were studied in [6] and subexponential slowdown asymptotics for ballistic RWRE similar to Theorem 1.4 were studied in [8]. The similarities to the current paper are greatest when the excited random walk has $\delta > 2$ and the RWRE is transient with positive speed and "nestling" (i.e., the environment has positive and negative drifts). In this case, the large deviation rate function for either model is zero on the interval $[0, v_0]$, where $v_0 = \lim_{n\to\infty} X_n/n$ is the limiting speed. Moreover, the polynomial rates of decay of the slowdown probabilities are related to the limiting distributions of the random walks in the same way. For instance, in either model if the slowdown probabilities decay like $n^{1-\alpha}$ with $\alpha \in (1, 2)$ then $n^{-1/\alpha}(X_n - nv_0)$ converges in distribution to an $\alpha$-stable random variable [14, 12].

An interesting difference in the rate functions for excited random walks and RWRE is that $I_X'(0) = 0$ in the present paper, while for transient RWRE the left and right derivatives of the rate function are not equal at the origin [6]. Since (in both models) $I_X$ is defined in terms of the large deviation rate function $I_T(t)$ for the hitting times $T_n/n$, this is related to the fact that $\inf_t I_T(t) = 0$ for excited random walks (see Lemma 4.1) while the corresponding rate function for the hitting times of RWRE is uniformly bounded away from $0$ if the walk is transient to the left.

### 1.3 Outline

The structure of the paper is as follows. In Section 2 we define the associated branching process with migration $V_i$, mention its relationship to the hitting times of the excited random walk, and prove a few basic properties about the process $V_i$. Then in Section 3 we prove a large deviation principle for the empirical mean of the process $V_i$ and prove some properties of the corresponding rate function. The large deviation principle for the empirical mean of the process $V_i$ is then used to deduce large deviation principles for $T_n/n$ and $X_n/n$ in Sections 4 and 5, respectively. Finally, in Section 6 we prove the subexponential rate of decay for slowdown probabilities.

## 2   A related branching process with random migration

In this section we recall how the hitting times $T_n$ of the excited random walk can be related to a branching process with migration. We will construct the related branching process with migration using the "coin tossing" construction that was given in [14]. Let a cookie environment $\omega = \{\omega_i(j)\}_{i\in\mathbb{Z}, j\geq 1}$ be fixed, and let $\{\xi_{i,j}\}_{i\in\mathbb{Z}, j\geq 1}$ be an independent family of Bernoulli random variables with $P(\xi_{i,j} = 1) = \omega_i(j)$. For $i$ fixed, we say that the $j$-th Bernoulli trial is a "success" if $\xi_{i,j} = 1$ and a "failure" otherwise. Then, let $F_m^{(i)}$ be the number of failures in the sequence $\{\xi_{i,j}\}_{j\geq 1}$ before the $m$-th success. That is,

$$F_m^{(i)} = \min\left\{\ell \geq 1 : \sum_{j=1}^{\ell} \xi_{i,j} = m\right\} - m.$$

Finally, we define the branching process with migration $\{V_i\}_{i\geq 1}$ by

$$V_0 = 0, \quad \text{and} \quad V_{i+1} = F_{V_i+1}^{(i)}, \quad \text{for } i \geq 0.$$

If the $\omega_i(j)$ were all equal to $1/2$ then the process $\{V_i\}$ would be a critical Galton-Watson branching process with one additional immigrant per generation. Allowing the first

$M$ cookie strengths at each site to be different than $1/2$ has the effect of making the migration effect more complicated (in particular, the migration in each generation is random and can depend on the current population size). We refer the interested reader to [1] for a more detailed description of the interpretation of $V_i$ as a branching process with migration.

In addition to the above branching process with migration, we will also need another branching process with a random initial population and one less migrant each generation. For any $n \geq 1$, let $V_0^{(n)} = V_n$ where $V_n$ is constructed as above and let $V_i^{(n)} = F_{V_{i-1}^{(n)}}^{(n+i-1)}$, where we let $F_0^{(i)} = 0$. Note that with this construction, we have that $V_i^{(n)} \leq V_{n+i}$ for all $i$. Moreover, while the Markov chain $V_i$ is irreducible, the lack of the extra migrant each generation makes $0$ an absorbing state for $V_i^{(n)}$.

The relevance of the processes $\{V_i\}_{i \geq 0}$ and $\{V_i^{(n)}\}_{i \geq 0}$ to the hitting times $T_n$ of the excited random walk is the following.

$$T_n \overset{\mathcal{D}}{=} n + 2\sum_{i=1}^{n} V_i + 2\sum_{i=1}^{\infty} V_i^{(n)}. \tag{2.1}$$

To explain this relation let $U_i^n = \#\{k \leq T_n \ : \ X_k = i, \ X_{k+1} = i-1\}$ be the number of times the random walk jumps from $i$ to $i-1$ before time $T_n$. Then, it is easy to see that $T_n = n + 2\sum_{i \leq n} U_i^n$ and (2.1) follows from the fact that

$$(U_n^n, U_{n-1}^n, \ldots U_1^n, U_0^n, U_{-1}^n, U_{-2}^n, \ldots) \overset{\mathcal{D}}{=} (V_1, V_2, \ldots, V_{n-1}, V_n, V_1^{(n)}, V_2^{(n)}, \ldots). \tag{2.2}$$

The details of the above joint equality in distribution can be found in [1] or [13].

**Remark 2.1.** *Technically, the relation* (2.1) *is proved in [1] and [13] only in the cases where $T_m < \infty$ with probability one. However, an examination of the proof shows that $P(T_n = k) = P(n + \sum_{i=1}^{n} V_i + 2\sum_{i=1}^{\infty} V_i^{(n)} = k)$ for any finite $k$ and so both sides of* (2.1) *are infinite with the same probability as well.*

## 2.1 Regeneration structure

We now define a sequence of regeneration times for the branching process $V_i$. Let $\sigma_0 = 0$ and for $k \geq 1$

$$\sigma_k = \inf\{i > \sigma_{k-1} : V_i = 0\}.$$

Also, for $k \geq 1$ let

$$W_k = \sum_{i=1}^{\sigma_k} V_i$$

be the total offspring of the branching process by the $k^{th}$ regeneration time. The tails of $\sigma_1$ and $W_1$ were analyzed in [13] in the case when $\delta > 0$.

**Lemma 2.2** (Theorems 2.1 and 2.2 in [13]). *If $\delta > 0$ then,*

$$P(\sigma_1 > x) \sim C_1 x^{-\delta} \quad and \quad P(W_1 > x) \sim C_2 x^{-\delta/2} \quad as \ x \to \infty. \tag{2.3}$$

Note that if the Markov chain $V_i$ is transient, then eventually $\sigma_k = W_k = \infty$ for all $k$ large enough. The following Lemma specifies the recurrence/transience properties of the Markov chain $V_i$.

**Lemma 2.3.** *The Markov chain $V_i$ is recurrent if and only if $\delta \geq 0$ and positive recurrent if and only if $\delta > 1$.*

*Proof.* The tail decay of $\sigma_1$ shows that $E[\sigma_1] < \infty$ if $\delta > 1$ and $E[\sigma_1] = \infty$ if $\delta \in (0, 1]$. Therefore, it is enough to show that $V_i$ is recurrent if and only if $\delta \geq 0$. This can be proven by an appeal to some previous results on branching proceses with migration as was done in [14]. A small difficulty arises in that the distribution of the migration that occurs before the generation of the $(i + 1)$-st generation depends on the population of $i$-th generation. However, this can be dealt with in the same manner as was done in [14]. To see this, let $\widehat{V}_i$ be defined by

$$\widehat{V}_0 = 0, \quad \text{and} \quad \widehat{V}_{i+1} = F^{(i)}_{(\widehat{V}_i+1)\vee M}.$$

Note that $V_i$ and $\widehat{V}_i$ have the same transition probabilities when starting from a site $k \geq M - 1$, and thus $V_i$ and $\widehat{V}_i$ are either both recurrent or both transient.

Next, let $Z_i = \widehat{V}_{i+1} - F^{(i)}_M$. We claim that $Z_i$ is recurrent if and only if $\widehat{V}_i$ is also recurrent. Since $0 \leq Z_i \leq \widehat{V}_{i+1}$, $Z_i$ is recurrent if $\widehat{V}_i$ is recurrent. To see the other implication, note that $Z_i = F^{(i)}_{(\widehat{V}_i+1)\vee M} - F^{(i)}_M$ is the number of failures in $\{\xi_{i,j}\}_{j\geq 1}$ between the $M$-th success and success number $(\widehat{V}_i + 1) \vee M$. Therefore, $Z_i$ is independent of $F^{(i)}_M$. Since $F^{(i)}_M$ is an i.i.d. sequence, then

$$\sum_{i\geq 0} P\left(\widehat{V}_{i+1} = 0\right) \geq \sum_{i\geq 0} P\left(Z_i = 0, F^{(i)}_M = 0\right) = P\left(F^{(0)}_M = 0\right) \sum_{i\geq 0} P(Z_i = 0),$$

and thus $\widehat{V}_i$ is recurrent if $Z_i$ is recurrent.

Finally, it can be shown that $Z_i$ is a branching process with migration where the migration component has mean $1 - \delta$ and the branching component has offspring distribution that is Geometric(1/2) (see Lemmas 16 and 17 in [14]). Then, previous results in the branching process with migration literature show that $Z_i$ is recurrent if and only if $\delta \geq 0$ (see Theorem A and Corollary 4 in [14] for a summary of these results). □

We close this section by noting that the above regeneration structure for the process $V_i$ can be used to give a representation for the limiting speed of the excited random walk. First note that, as was shown in [1], the representation (2.1) can be used to show that when $\delta > 1$,

$$\frac{1}{v_0} = \lim_{n\to\infty} \frac{T_n}{n} = 1 + 2 \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} V_i.$$

To compute the last limit above, first note that $\{(W_k - W_{k-1}, \sigma_k - \sigma_{k-1})\}_{k\geq 1}$ is an i.i.d. sequence and that the tail decay of $\sigma_1$ given in Theorem 2.1 of [13] implies that $E[\sigma_1] < \infty$ whenever $\delta > 1$. Let $k(n)$ be defined by $\sigma_{k(n)-1} < n \leq \sigma_{k(n)}$. A standard renewal theory argument implies that

$$\lim_{n\to\infty} \frac{k(n)}{n} = \frac{1}{E[\sigma_1]}.$$

Since $W_{k(n)-1} \leq \sum_{i=1}^{n} V_n \leq W_{k(n)}$ and $\lim_{k\to\infty} W_k/k = E[W_1]$, this implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} V_i = \frac{E[W_1]}{E[\sigma_1]}.$$

Therefore, we obtain the following formula for the limiting speed of transient excited random walks.

**Lemma 2.4.** *If $\delta > 1$, then*

$$v_0 = \frac{\mathbb{E}[\sigma_1]}{\mathbb{E}[\sigma_1 + 2W_1]}. \tag{2.4}$$

**Remark 2.5.** *The tail decay of $W_1$ in (2.3) implies that $E[W_1] = \infty$ when $\delta \in (1,2]$. However, the limiting speed $v_0 = 0$ when $\delta \in (1,2]$ so that the equality (2.4) still holds in this case.*

## 3  Large Deviations for the Branching Process

In this section we discuss the large deviations of $n^{-1} \sum_{i=1}^{n} V_i$. Let

$$\Lambda_{W,\sigma}(\lambda, \eta) = \log \mathbb{E}\left[ e^{\lambda W_1 + \eta \sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}} \right], \qquad \lambda, \eta \in \mathbb{R}, \tag{3.1}$$

be the logarithmic moment generating function of $(W_1, \sigma_1)$, and let

$$\Lambda_V(\lambda) = -\sup\{\eta : \Lambda_{W,\sigma}(\lambda, \eta) \leq 0\} \quad \text{and} \quad I_V(x) = \sup_{\lambda} \lambda x - \Lambda_V(\lambda). \tag{3.2}$$

The relevance of these functions is seen by the following Theorem, which is a direct application of a more general result of Nummelin and Ney (see remark (ii) at the bottom of page 594 in [18]).

**Theorem 3.1.** *Let $I_V(x)$ be defined as in (3.2). Then,*

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left( \frac{1}{n} \sum_{i=1}^{n} V_i \in G, V_n = j \right) \geq -\inf_{x \in G} I_V(x),$$

*for all open $G$ and any $j \geq 0$, and*

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left( \frac{1}{n} \sum_{i=1}^{n} V_i \in F, V_n = j \right) \leq -\inf_{x \in F} I_V(x),$$

*for all closed $F$ and any $j \geq 0$.*

In order to obtain large deviation results for the related excited random walk, it will also be necessary to obtain large deviation asymptotics of $n^{-1} \sum_{i=1}^{n} V_i$ without the added condition on the value of $V_n$.

**Theorem 3.2.** *Let $I_V(x)$ be defined as in (3.2). Then, $n^{-1} \sum_{i=1}^{n} V_i$ satisfies a large deviation principle with rate function $I_V(x)$. That is,*

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left( \frac{1}{n} \sum_{i=1}^{n} V_i \in G \right) \geq -\inf_{x \in G} I_V(x),$$

*for all open $G$, and*

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left( \frac{1}{n} \sum_{i=1}^{n} V_i \in F \right) \leq -\inf_{x \in F} I_V(x), \tag{3.3}$$

*for all closed $F$.*

**Remark 3.3.** *There are many results in the large deviations literature that imply a large deviation principle for the empirical mean of a Markov chain. However, we were not able to find a suitable theorem that implied Theorem 3.2. Some of the existing results required some sort of fast mixing of the Markov chain [5, 9], but the Markov chain $\{V_i\}_{i \geq 0}$ mixes very slowly since if $V_0$ is large it typically takes a long time to return to $0$ (on the order of $\mathcal{O}(V_0)$ steps). Moreover, it is very important that the rate functions are the same in Theorems 3.1 and 3.2, and many of the results for the large deviations for the empirical mean of a Markov chain formulate the rate function in terms of the spectral radius of an operator [7] instead of in terms of logarithmic moment generating functions as in (3.1) and (3.2).*

*Proof.* Obviously the lower bound in Theorem 3.2 follows from the corresponding lower bound in (3.1), and so it is enough to prove the upper bound only. Our proof will use the following facts about the functions $\Lambda_V$ and $I_V$.

(i)   $\Lambda_V(\lambda)$ is convex and continuous on $(-\infty, 0]$ and $\Lambda_V(\lambda) = \infty$ for all $\lambda > 0$. Therefore, $I_V(x) = \sup_{\lambda < 0} (\lambda x - \Lambda_V(\lambda))$.

(ii)   $I_V(x)$ is a convex, non-increasing function of $x$, and $\lim_{x \to \infty} I_V(x) = \inf_x I_V(x) = 0$.

These properties and more will be shown in Section 3.1 below where we give a qualitative description of the rate function $I_V$. By property (ii) above, it will be enough to prove the large deviation upper bound for closed sets of the form $F = (-\infty, x]$. That is, we need only to show that

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left(\sum_{i=1}^n V_i \le xn\right) \le -I_V(x), \quad \forall x < \infty. \tag{3.4}$$

This will follow from

$$\limsup_{n \to \infty} \frac{1}{n} \log E[e^{\lambda \sum_{i=1}^n V_i}] \le \Lambda_V(\lambda), \quad \forall \lambda < 0. \tag{3.5}$$

Indeed, combining (3.5) with the usual Chebyshev upper bound for large deviations gives that for any $x < \infty$ and $\lambda < 0$,

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left(\sum_{i=1}^n V_i \le xn\right) \le \limsup_{n \to \infty} \frac{1}{n} \log(e^{-\lambda xn} E[e^{\lambda \sum_{i=1}^n V_i}]) \le -\lambda x + \Lambda_V(\lambda).$$

Optimizing over $\lambda < 0$ and using property (i) above proves (3.4).

It remains still to prove (3.5). By decomposing according to the time of the last regeneration before $n$ we obtain

$$E[e^{\lambda \sum_{i=1}^n V_i - \Lambda_V(\lambda)n}]$$

$$= E[e^{\lambda \sum_{i=1}^n V_i - \Lambda_V(\lambda)n} \mathbb{1}_{\{\sigma_1 > n\}}] + \sum_{m=1}^n \sum_{t=0}^{n-1} E[e^{\lambda \sum_{i=1}^n V_i - \Lambda_V(\lambda)n} \mathbb{1}_{\{\sigma_m \le n < \sigma_{m+1}, \, n - \sigma_m = t\}}]$$

$$= E[e^{\lambda \sum_{i=1}^n V_i - \Lambda_V(\lambda)n} \mathbb{1}_{\{\sigma_1 > n\}}]$$

$$\quad + \sum_{m=1}^n \sum_{t=0}^{n-1} E[e^{\lambda W_m - \Lambda_V(\lambda)\sigma_m} \mathbb{1}_{\{\sigma_m = n-t\}}] E[e^{\lambda \sum_{i=1}^t V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}]$$

$$\le E[e^{\lambda \sum_{i=1}^n V_i - \Lambda_V(\lambda)n} \mathbb{1}_{\{\sigma_1 > n\}}] \tag{3.6}$$

$$\quad + \left(\sum_{m=1}^n E[e^{\lambda W_m - \Lambda_V(\lambda)\sigma_m} \mathbb{1}_{\{\sigma_m < \infty\}}]\right) \left(\sum_{t=0}^{n-1} E[e^{\lambda \sum_{i=1}^t V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}]\right), \tag{3.7}$$

where we used the Markov property in the second equality. The definition of $\Lambda_V$ and the monotone convergence theorem imply that $\Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda)) \le 0$. Therefore,

$$\sum_{m=1}^n E[e^{\lambda W_m - \Lambda_V(\lambda)\sigma_m} \mathbb{1}_{\{\sigma_m < \infty\}}] = \sum_{m=1}^n e^{m \Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda))} \le n.$$

To bound the second sum in (3.7) we need the following lemma, whose proof we postpone for now.

**Lemma 3.4.** *For any $\lambda < 0$,*

$$\sup_{t \geq 0} E[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}] < \infty.$$

Lemma 3.4 implies the expectation (3.6) is uniformly bounded in $n$ and that the second sum in (3.7) grows at most linearly in $n$. Since the first sum in (3.7) also grows linearly in $n$ this implies that

$$\limsup_{n \to \infty} \frac{1}{n} \log E[e^{\lambda \sum_{i=1}^{n} V_i - \Lambda_V(\lambda)n}] \leq 0, \quad \forall \lambda < 0,$$

which is obviously equivalent to (3.5). It remains only to give the proof of Lemma 3.4.

*Proof of Lemma 3.4.* First, note that

$$
\begin{aligned}
1 &\geq E[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] \\
&\geq E[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{t < \sigma_1 < \infty\}}] \\
&= E\left[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}} e^{\lambda \sum_{i=t+1}^{\sigma_1} V_i - \Lambda_V(\lambda)(\sigma_1 - t)} \mathbb{1}_{\{\sigma_1 < \infty\}}\right] \\
&= E\left[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}} E^{V_t}\left[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}\right]\right],
\end{aligned}
\tag{3.8}
$$

where in the last equality we use the notation $E^m$ for the expectation with respect to the law of the Markov chain $V_i$ conditioned on $V_0 = m$. Since $V_i$ is an irreducible Markov chain and $E[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] \leq 1$, then the inner expectation in (3.8) is finite for any value of $V_t$ and can be uniformly bounded below if $V_t$ is restricted to a finite set. Thus, for any $K < \infty$,

$$
\begin{aligned}
&E\left[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t, \, V_t \leq K\}}\right] \\
&\leq \left(\inf_{m \in [1,K]} E^m[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}]\right)^{-1} E[e^{\lambda W_1 - \Lambda_V(\lambda)\sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}].
\end{aligned}
\tag{3.9}
$$

Let $C_{K,\lambda} < \infty$ be defined to be the right side of the inequality above.

Note that the upper bound (3.9) does not depend on $t$. The key to finishing the proof of Lemma 3.4 is using the upper bound (3.9) in an iterative way. For any $t \geq 1$,

$$
\begin{aligned}
E[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}] &\leq C_{K,\lambda} + E[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t, \, V_t > K\}}] \\
&\leq C_{K,\lambda} + e^{\lambda K - \Lambda_V(\lambda)} E[e^{\lambda \sum_{i=1}^{t-1} V_i - \Lambda_V(\lambda)(t-1)} \mathbb{1}_{\{\sigma_1 > t-1\}}],
\end{aligned}
$$

where in the last inequality we used that $\{\sigma_1 > t, \, V_t > K\} = \{\sigma_1 > t-1, \, V_t > K\}$. Iterating the above bound implies that

$$E[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}] \leq C_{K,\lambda} \sum_{l=0}^{t-1} e^{l(\lambda K - \Lambda_V(\lambda))} + e^{t(\lambda K - \Lambda_V(\lambda))}.$$

By choosing $K > \Lambda_V(\lambda)/\lambda$ so that $e^{\lambda K - \Lambda_V(\lambda)} < 1$, we thus obtain that

$$\sup_{t \geq 0} E[e^{\lambda \sum_{i=1}^{t} V_i - \Lambda_V(\lambda)t} \mathbb{1}_{\{\sigma_1 > t\}}] \leq \frac{C_{K,\lambda}}{1 - e^{K\lambda - \Lambda_V(\lambda)}} + 1 < \infty.$$

$\square$

$\square$

### 3.1 Properties of the rate function $I_V$

We now turn our attention to a qualitative description of the rate function $I_V$. Since $I_V$ is defined as the Legendre dual of $\Lambda_V$, these properties will in turn follow from an understanding of $\Lambda_V$ (and also $\Lambda_{W,\sigma}$). We begin with some very basic properties of $\Lambda_V$ and the corresponding properties of $I_V$.

**Lemma 3.5.** *$\Lambda_V(\lambda)$ is non-decreasing, convex, and left-continuous as a function of $\lambda$. Moreover,*

(i) *$\Lambda_V(\lambda) \in (\log \mathbf{E}[\omega_0(1)], 0)$ for all $\lambda < 0$, and $\lim_{\lambda \to -\infty} \Lambda_V(\lambda) = \log \mathbf{E}[\omega_0(1)]$.*
(ii) *$\Lambda_V(\lambda) = \infty$ if $\lambda > 0$.*

*Proof.* Recall the definitions of $\Lambda_{W,\sigma}$ and $\Lambda_V$ in (3.1) and (3.2), respectively. The fact that $\Lambda_V(\lambda)$ is non-decreasing follows from the fact that $\Lambda_{W,\sigma}(\lambda_1, \eta) \leq \Lambda_{W,\sigma}(\lambda_2, \eta)$ for any $\lambda_1 < \lambda_2$. Since $\Lambda_{W,\sigma}$ is the logarithmic generating function of the joint random variables $(W_1, \sigma_1)$, then $\Lambda_{W,\sigma}(\lambda, \eta)$ is a convex function of $(\lambda, \eta)$ (and strictly convex on $\{(\lambda, \eta) : \Lambda_{W,\sigma}(\lambda, \eta) < \infty\}$). The convexity of $\Lambda_V$ as a function of $\lambda$ then follows easily from the convexity of $\Lambda_{W,\sigma}(\lambda, \eta)$ and the definition of $\Lambda_V$. Also, left-continuity follows from the definition of $\Lambda_V$ and the fact that $\lim_{\lambda \to \lambda'} \Lambda_{W,\sigma}(\lambda, \eta) = \Lambda_{W,\sigma}(\lambda', \eta)$ by the monotone convergence theorem.

Since $W_1 \geq 0$, $\Lambda_{W,\sigma}(\lambda, 0) = \log E[e^{\lambda W_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] < 0$ for all $\lambda < 0$. On the other hand, since $W_1 \geq \sigma_1 - 1$ it follows that $\Lambda_{W,\sigma}(\lambda, -\lambda) \leq -\lambda < \infty$ for all $\lambda \leq 0$. Then the continuity of $\Lambda_{W,\sigma}$ and the definition of $\Lambda_V(\lambda)$ imply that $\Lambda_V(\lambda) < 0$ for all $\lambda < 0$. Additionally,

$$E[e^{\lambda W_1 + \eta \sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] > e^\eta P(\sigma_1 = 1) = e^\eta \mathbf{E}[\omega_0(1)],$$

which implies that $\Lambda_{W,\sigma}(\lambda, -\log \mathbf{E}[\omega_0(1)]) > 0$ for all $\lambda < 0$. Thus, $\Lambda_V(\lambda) > \log \mathbf{E}[\omega_0(1)]$ for all $\lambda < 0$. To prove the second part of property (i), note that $\sigma_1 \leq W_1 + 1$ implies that for $\eta \geq 0$,

$$\lim_{\lambda \to -\infty} E[e^{\lambda W_1 + \eta \sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] \leq \lim_{\lambda \to -\infty} E[e^{(\lambda + \eta)W_1 + \eta}] = e^\eta P(W_1 = 0) = e^\eta \mathbf{E}[\omega_0(1)], \quad (3.10)$$

where the second to last equality follows from the bounded convergence theorem. From (3.10) and the definition of $\Lambda_V$, it follows that $\lim_{\lambda \to -\infty} \Lambda_V(\lambda) \leq \log \mathbf{E}[\omega_0(1)]$. Combining this with the first part of property (i) implies the second part of property (i).

To show that $\Lambda_V(\lambda) = \infty$ for $\lambda > 0$ it is actually easiest to refer back to the excited random walk. Recall the naive strategy for slowdowns of the excited random walk in Example 1.3. We can modify the strategy slightly by not only consuming all cookies in $[0, n^{1/3})$ and then staying in the interval for $n$ steps, but also requiring that the random walk then exits the interval on the right. This event still has a probability bounded below by $Ce^{-cn^{1/3}}$. Examining the branching process corresponding to the excited random walk we see that the event for the random walk described above implies that $U_i^N \geq 1$ for all $i \in [1, N-1]$, $U_0^N = 0$ and $\sum_{i=1}^N U_i^N > n/2$, where $N = \lceil n^{1/3} \rceil$. Then, using (2.2) we obtain that that $P(W_1 > n/2, \sigma_1 = \lceil n^{1/3} \rceil) \geq Ce^{-cn^{1/3}}$ for all $n \geq 1$ which implies that

$$E[e^{\lambda W_1 + \eta \sigma_1} \mathbb{1}_{\{\sigma_1 < \infty\}}] \geq e^{\lambda n/2 + \eta n^{1/3}} P(W_1 > n/2, \sigma_1 = \lceil n^{1/3} \rceil) \geq Ce^{\lambda n/2 + \eta n^{1/3} - cn^{1/3}},$$

for any $\lambda > 0$ and $\eta < 0$. Since this lower bound can be made arbitrarily large by taking $n \to \infty$, this shows that $\Lambda_{W,\sigma}(\lambda, \eta) = \infty$ for any $\lambda > 0$ and $\eta < 0$, and thus $\Lambda_V(\lambda) = \infty$ for all $\lambda > 0$. $\qquad\square$

We would like to say that $\Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda)) = 0$. However, in order to be able to conclude this is true, we need to show that $\Lambda_{W,\sigma}(\lambda, \eta) \in [0, \infty)$ for some $\eta$. The next series of lemmas gives some conditions where we can conclude this is true.

**Lemma 3.6.** *If $\lambda \leq \log \mathbf{E}[\omega_0(1)]$, then*

$$\Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda)) = 0. \tag{3.11}$$

*Moreover, $\Lambda_V(\lambda)$ is strictly convex and analytic on $(-\infty, \log \mathbf{E}[\omega_0(1)])$.*

*Proof.* Since $W_1 \geq \sigma_1 - 1$ we have that for $\lambda \leq 0$,

$$\Lambda_{W,\sigma}(\lambda, -\lambda) = \log E[e^{\lambda(W_1 - \sigma_1)}\mathbb{1}_{\{\sigma_1 < \infty\}}] \leq -\lambda.$$

Therefore, $\Lambda_{W,\sigma}(\lambda, \eta) < \infty$ for all $\lambda < 0$ and $\eta \leq -\lambda$. On the other hand, it was shown above that $\Lambda_{W,\sigma}(\lambda, 0) < 0$ and $\Lambda_{W,\sigma}(\lambda, -\log \mathbf{E}[\omega_0(1)]) > 0$ when $\lambda < 0$. Since $\Lambda_{W,\sigma}(\lambda, \eta)$ is monotone increasing and continuous in $\eta$ this implies that $\Lambda_{W,\sigma}(\lambda, \eta) = 0$ has a unique solution $\eta \in [0, -\log \mathbf{E}[\omega_0(1)]]$ when $\lambda \leq \log \mathbf{E}[\omega_0(1)]$. By the definition of $\Lambda_V$ and the fact that $\Lambda_{W,\sigma}(\lambda, \eta)$ is strictly increasing in $\eta$, this must be $\eta = -\Lambda_V(\lambda)$.

Let $\mathcal{D}_{W,\sigma} = \{(\lambda, \eta) : \Lambda_{W,\sigma}(\lambda, \eta) < \infty\}$. The above argument shows not only that (3.11) holds but also that $(\lambda, -\Lambda_V(\lambda))$ is in the interior of $\mathcal{D}_{W,\sigma}$ when $\lambda \leq \log \mathbf{E}[\omega_0(1)]$. Since $\Lambda_{W,\sigma}$ is analytic on $\mathcal{D}_{W,\sigma}$, the implicit function theorem implies that $\Lambda_V(\lambda)$ is analytic on $(-\infty, \log \mathbf{E}[\omega_0(1)])$. Finally, combining (3.11) with the fact that $\Lambda_{W,\sigma}$ is strictly convex on $\mathcal{D}_{W,\sigma}$ implies that $\Lambda_V(\lambda)$ is strictly convex on $(-\infty, \log \mathbf{E}[\omega_0(1)])$. $\qquad\square$

**Lemma 3.7.** *For every $m < \infty$, there exists a $\lambda_0 = \lambda_0(m) < 0$ such that $\Lambda_{W,\sigma}(\lambda, -\lambda m) < \infty$ for all $\lambda \in (\lambda_0, 0)$.*

*Proof.* We need to show that $E[e^{\lambda W_1 - \lambda m \sigma_1}\mathbb{1}_{\{\sigma_1 < \infty\}}] = E[e^{\lambda \sum_{i=1}^{\sigma_1}(V_i - m)}\mathbb{1}_{\{\sigma_1 < \infty\}}] < \infty$ for $\lambda$ negative and sufficiently close to zero. Since $\lambda < 0$ we need to bound the sum in the exponent from below. Note that all the terms in the sum except the last one are larger than $-(m-1)$ and that the terms are non-negative if $V_i \geq m$. Therefore, letting $N_m = \#\{1 \leq i \leq \sigma_1 : V_i < m\}$ we obtain that

$$E[e^{\lambda W_1 - \lambda m \sigma_1}\mathbb{1}_{\{\sigma_1 < \infty\}}] \leq E[e^{-\lambda(m-1)N_m}].$$

To show that this last expectation is finite for $\lambda$ close to zero, we need to show that $N_m$ has exponential tails. To this end, note that the event $\{N_m > n\}$ implies that the first $n$ times that the process $V_i < m$, the following step is not to 0. Thus,

$$P(N_m > n) \leq \left(\max_{k < m} P(V_1 \neq 0 \,|\, V_0 = k)\right)^n = P(V_1 \neq 0 \,|\, V_0 = \lceil m \rceil - 1)^n.$$

Therefore, the statement of the Lemma holds with

$$\lambda_0(m) = \frac{1}{m-1}\log P(V_1 \neq 0 \,|\, V_0 = \lceil m \rceil - 1).$$

$\qquad\square$

**Corollary 3.8.** *If $\delta > 2$ (so that $E[W_1], E[\sigma_1] < \infty$), then there exists a $\lambda_1 < 0$ such that on the interval $(\lambda_1, 0)$*

  *(i)* $\Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda)) = 0$.
 *(ii)* $\Lambda_V(\lambda)$ *is analytic and strictly convex as a function of $\lambda$.*
*(iii)* $\lim_{\lambda \to 0^+} \Lambda_V'(\lambda) = E[W_1]/E[\sigma_1] =: m_0$.

*Proof.* Let $\mathcal{D}_{W,\sigma} = \{(\lambda, \eta) : \Lambda_{W,\sigma}(\lambda, \eta) < \infty\}$ be the domain where $\Lambda_{W,\sigma}$ is finite, and let $D_{W,\sigma}^\circ$ be the interior of $D_{W,\sigma}$. Define $m_0 = E[W_1]/E[\sigma_1]$. Then, if $0 > \lambda > \lambda_0(m)m/m_0$ for some $m > m_0$, it follows that

$$\Lambda_{W,\sigma}(\lambda, -\lambda m_0) \leq \Lambda_{W,\sigma}\left(\frac{\lambda m_0}{m}, -\lambda m_0\right) < \infty,$$

where the first inequality follows from the monotonicity of $\Lambda_{W,\sigma}$ and the last inequality follows from Lemma 3.7. Thus, $(\lambda, -\lambda m_0) \in \mathcal{D}_{W,\sigma}^\circ$ if $\lambda \in (\lambda_1, 0)$ with

$$\lambda_1 = \inf_{m > m_0} \frac{\lambda_0(m)m}{m_0}.$$

Since $\Lambda_{W,\sigma}$ is analytic and strictly convex in $D_{W,\sigma}^\circ$, the function $g(\lambda) = \Lambda_{W,\sigma}(\lambda, -m_0\lambda)$ is strictly convex and analytic on the interval $(\lambda_1, 0)$. In particular, $g$ is differentiable and

$$g'(\lambda) = \frac{d}{d\lambda} \log E\left[e^{\lambda(W_1 - m_0\sigma_1)}\right] = \frac{E\left[(W_1 - m_0\sigma_1)e^{\lambda(W_1 - m_0\sigma_1)}\right]}{E\left[e^{\lambda(W_1 - m_0\sigma_1)}\right]}.$$

Since $g$ is strictly convex,

$$g'(\lambda) < \lim_{\lambda \to 0^-} g'(\lambda) = E[W_1 - m_0\sigma_1] = 0, \qquad \forall \lambda \in (\lambda_1, 0).$$

Therefore, $g(\lambda)$ is strictly decreasing on $(\lambda_1, 0)$. Since, $\lim_{\lambda \to 0^-} g(\lambda) = g(0) = 0$ we obtain that $g(\lambda) = \Lambda_{W,\sigma}(\lambda, -m_0\lambda) > 0$ for $\lambda \in (\lambda_1, 0)$. Thus, for every $\lambda \in (\lambda_1, 0)$ there exists an $\eta \in (0, -m_0\lambda)$ such that $\Lambda_{W,\sigma}(\lambda, \eta) = 0$, and the definition of $\Lambda_V$ implies that $\eta = -\Lambda_V(\lambda)$. We have shown that $\Lambda_{W,\sigma}(\lambda, -\Lambda_V(\lambda)) = 0$ and $(\lambda, -\Lambda_V(\lambda)) \in \mathcal{D}_{W,\sigma}^\circ$ for all $\lambda \in (\lambda_1, 0)$. As was the case in the proof of Lemma 3.6 these facts imply that $\Lambda_V(\lambda)$ is analytic and strictly convex on $(\lambda_1, 0)$.

To show that $\lim_{\lambda \to 0^-} \Lambda_V'(\lambda) = m_0$, first note that as was shown above $\Lambda_V(\lambda) > m_0\lambda$ for $\lambda < 0$. For $m < m_0$ define $g_m(\lambda) = \Lambda_{W,\sigma}(\lambda, -m\lambda)$. For $\lambda$ close enough to 0 we have that $g_m(\lambda)$ is strictly convex and analytic, and that

$$g_m'(\lambda) = \frac{d}{d\lambda} \log E\left[e^{\lambda(W_1 - m\sigma_1)}\right] = \frac{E\left[(W_1 - m\sigma_1)e^{\lambda(W_1 - m\sigma_1)}\right]}{E\left[e^{\lambda(W_1 - m\sigma_1)}\right]}.$$

Therefore, $\lim_{\lambda \to 0^-} g_m'(\lambda) = E[W_1] - mE[\sigma_1] > 0$, and thus there exists a $\lambda_2 = \lambda_2(m) < 0$ such that $g_m(\lambda) = \Lambda_{W,\sigma}(\lambda, -m\lambda) < 0$ for $\lambda \in (\lambda_2, 0)$. This implies that $m_0\lambda < \Lambda_V(\lambda) < m\lambda$ for all $\lambda \in (\lambda_2, 0)$, and thus $\lim_{\lambda \to 0^-} \Lambda_V'(\lambda) \in [m, m_0]$. The proof is finished by noting that this is true for any $m < m_0$. $\qquad\square$

We are now ready to deduce some properties of the rate function $I_V$.

**Lemma 3.9.** $\inf_x I_V(x) = 0$.

*Proof.* Since $I_V$ is the Legendre transform of $\Lambda_V$ and $\Lambda_V$ is lower-semicontinuous, then it follows that $\inf_x I_V(x) = -\Lambda_V(0)$. If $\delta \geq 0$, then $\Lambda_{W,\sigma}(0,0) = \log P(\sigma_1 < \infty) = 0$ (by Lemma 2.3) and thus $\Lambda_V(0) = 0$ when $\delta \geq 0$.

If $\delta < 0$, then $\Lambda_{W,\sigma}(0,0) = \log P(\sigma_1 < \infty) < 0$ and so we can only conclude a priori[1] that $\Lambda_V(0) \leq 0$. Instead we will prove $\inf_x I_V(x) = 0$ in a completely different manner. First note that letting $F = G = \mathbb{R}$ in Theorem 3.2 implies that

$$\lim_{n \to \infty} \frac{1}{n} \log P(V_n = 0) = -\inf_x I_V(x).$$

Therefore, we need to show that $P(V_n = 0)$ does not decay exponentially fast in $n$. The explanation of the representation (2.1) implies that $P(T_n < T_{-1}) = P(U_0^n = 0) \leq P(V_n = 0)$, and thus we are reduced to showing that $P(T_n < T_{-1})$ does not decay exponentially fast in $n$. In fact, we claim that there exists a constant $C > 0$ such that

$$P(T_n < T_{-1}) \geq Cn^{-M-1} \tag{3.12}$$

---

[1]Note that we cannot conclude that $\Lambda_V(0) < 0$ since we do not know if $\Lambda_{W,\sigma}(0,\eta) < \infty$ for some $\eta > 0$. In fact since $\inf_x I_V(x) = 0$ if and only if $\Lambda_V(0) = 0$, the proof of the lemma shows indirectly that $\Lambda_{W,\sigma}(0,\eta) = \infty$ for all $\eta > 0$ when $\delta < 0$.

To see this, suppose that the first $2M + 1$ steps of the random walk alternate between 0 and 1. The probability of this happening is

$$P(X_{2i} = 0, \, X_{2i+1} = 1, \, i = 1, 2, \ldots M) = \mathbf{E}\left[\prod_{j=1}^{M+1} \omega_0(j) \prod_{j=1}^{M} (1 - \omega_1(j))\right]$$

$$= \frac{1}{2}\mathbf{E}\left[\prod_{j=1}^{M} \omega_0(j)\right] \mathbf{E}\left[\prod_{j=1}^{M} (1 - \omega_0(j))\right] > 0.$$

At this point the random walker has consumed all the "cookies" at the sites $0$ and $1$. Therefore, by a simple symmetric random walk computation, the probability that the random walk from this point hits $x = 2$ before $x = -1$ is $2/3$. Since $\delta < 0$ the random walk will eventually return from $x = 2$ to $x = 1$ again, and then the probability that the random walk again jumps $M$ more times from $x = 1$ to $x = 2$ without hitting $x = -1$ is $(2/3)^M$. After jumping from $x = 1$ to $x = 2$ a total of $M + 1$ times there are no longer any cookies at $x = 2$ either, and thus the probability that the random walk now jumps $M + 1$ times from $x = 2$ to $x = 3$ without visiting $x = -1$ is $(3/4)^{M+1}$. We continue this process at successive sites to the right until the random walk makes $M + 1$ jumps from $x = n - 2$ to $x = n - 1$ without hitting $x = -1$ (which happens with probability $((n - 1)/n)^{M+1}$). Upon this last jump to $x = n - 1$ the random walk has consumed all cookies at $x = n - 1$ and so the probability that the next step is to the right is $1/2$. Putting together the above information we obtain the lower bound

$$P(T_n < T_{-1}) \geq \left(\frac{1}{2}\mathbf{E}\left[\prod_{j=1}^{M} \omega_0(j)\right] \mathbf{E}\left[\prod_{j=1}^{M} (1 - \omega_0(j))\right]\right)\left(\frac{2}{3}\frac{3}{4}\cdots\frac{n-1}{n}\right)^{M+1}\frac{1}{2}.$$

This completes the proof of (3.12), and thus $\inf_x I_V(x) = 0$ when $\delta < 0$. $\qquad\square$

**Lemma 3.10.** *The function $I_V(x)$ is convex, non-increasing, and continuous on $[0, \infty)$. Moreover,*

(i) *There exists an $m_2 > 0$ such that $I_V(x)$ is strictly convex and analytic on $(0, m_2)$.*

(ii) *$I_V(0) = -\log \mathbf{E}[\omega_0(1)]$ and $\lim_{x \to 0^+} I'_V(x) = -\infty$.*

(iii) *If $\delta > 2$ then there exists an $m_1 < m_0 = E[W_1]/E[\sigma_1]$ such that $I_V(x)$ is strictly convex and analytic on $(m_1, m_0)$, $I_V(x) = 0$ for $x \geq m_0$, and $\lim_{x \to m_0^-} I'_V(x) = 0$ so that $I_V$ is continuously differentiable on $(m_1, \infty)$.*

(iv) *If $\delta \leq 2$ then $I_V(x) > 0$ for all $x < \infty$.*

*Proof.* Since $I_V$ is the Legendre transform of $\Lambda_V$, $I_V(x)$ is lower-semicontinuous and convex as a function in $x$. It follows easily from Lemma 3.5 and the definition of $I_V$ that $I_V(x) < \infty$ if and only if $x \in [0, \infty)$, and since $I_V$ is convex and lower-semicontinuous this shows that $I_V$ is continuous on $[0, \infty)$. The fact that $I_V(x)$ is non-increasing follows from the fact that $\Lambda_V(\lambda) = \infty$ for any $\lambda > 0$. Indeed, if $x_1 \leq x_2$ then

$$I_V(x_1) = \sup_{\lambda \leq 0}\{\lambda x_1 - \Lambda_V(\lambda)\} \geq \sup_{\lambda \leq 0}\{\lambda x_2 - \Lambda_V(\lambda)\} = I_V(x_2).$$

Next, recall from Lemma 3.6 that on $(-\infty, \log \mathbf{E}[\omega_0(1)])$ the function $\Lambda_V(\lambda)$ is strictly convex and analytic and let $m_2 = \lim_{\lambda \to \log \mathbf{E}[\omega_0(1)]^-} \Lambda'_V(\lambda)$. The fact that $\Lambda_V(\lambda)$ is non-decreasing and uniformly bounded below also implies that $\lim_{\lambda \to -\infty} \Lambda'_V(\lambda) = 0$. Therefore, for every $x \in (0, m_2)$ there exists a unique $\lambda = \lambda(x)$ such that $\Lambda'_V(\lambda(x)) = x$ and so

$$I_V(x) = \lambda(x)x - \Lambda_V(\lambda(x)) \quad \text{for } x \in (0, m_2). \tag{3.13}$$

Since $\Lambda_V(\lambda)$ is analytic on $(-\infty, \log \mathbf{E}[\omega_0(1)])$ the inverse function theorem implies that $\lambda(x)$ is analytic on $(0, m_2)$ and thus (3.13) implies that $I_V(x)$ is analytic on $(0, m_2)$ as well. To see that $I_V(x)$ is strictly convex on $(0, m_2)$, we differentiate (3.13) with respect to $x$ and use the fact that $\Lambda'_V(\lambda(x)) = x$ for $x \in (0, m_2)$ to obtain

$$I'_V(x) = \lambda(x), \quad \text{for } x \in (0, m_2). \tag{3.14}$$

Since $\lambda(x)$ is strictly increasing on $(0, m_2)$, it follows that $I_V$ is strictly convex on $(0, m_2)$. Moreover, (3.14) implies that $\lim_{x \to 0+} I'_V(x) = \lim_{x \to 0+} \lambda(x) = -\infty$ and Lemma 3.5 (i) implies that that $I_V(0) = -\inf_\lambda \Lambda_V(\lambda) = -\log \mathbf{E}[\omega_0(1)]$.

When $\delta > 2$, Lemma 3.8 implies that $\Lambda_V(\lambda)$ is analytic and strictly convex on $(\lambda_1, 0)$. Let $m_1 = \lim_{\lambda \to \lambda_1^+} \Lambda'_V(\lambda)$ and recall that $\lim_{\lambda \to 0^-} \Lambda'_V(\lambda) = m_0 = E[W_1]/E[\sigma_1]$. Then the same argument as above shows that $I_V(x)$ is strictly convex and analytic on $(m_1, m_0)$ and that $\lim_{x \to m_0^-} I'_V(x) = 0$. Now, since $\Lambda'_V(\lambda)$ increases to $m_0$ as $\lambda \to 0^-$ and $\Lambda_V(0) = 0$, then $\Lambda_V(\lambda) \geq m_0 \lambda$ for all $\lambda \leq 0$. Therefore

$$I_V(x) = \sup_{\lambda \leq 0} \{\lambda x - \Lambda_V(\lambda)\} \leq \sup_{\lambda \leq 0} \{\lambda m_0 - \Lambda_V(\lambda)\} = 0, \quad \text{for all } x \geq m_0,$$

where the first equality follows from the fact that $\Lambda_V(\lambda) = \infty$ if $\lambda > 0$. However, since $I_V(x) \geq -\Lambda_V(0) = 0$ it must be that $I_V(x) = 0$ for $x \geq m_0$.

It remains only to show that $I_V(x) > 0$ for all $x$ when $\delta \leq 2$. We will divide the proof into two cases: $\delta \in (1, 2]$ and $\delta \leq 1$.

**Case I:** $\delta \in (1, 2]$.

For any $m < \infty$ let $g_m(\lambda) = \Lambda_{W,\sigma}(\lambda, -m\lambda)$. Then, as in the proof of Corollary 3.8, $g_m(\lambda)$ is analytic and strictly convex for $\lambda < 0$ close enough to zero. Moreover,

$$\lim_{\lambda \to 0^-} g'_m(\lambda) = E[W_1 - m\sigma_1] = \infty,$$

where the last equality holds since the tail decay of $W_1$ and $\sigma_1$ in (2.3) implies that $E[W_1] = \infty$ and $E[\sigma_1] < \infty$ when $\delta \in (1, 2]$. Since $g_m(0) = \Lambda_{W,\sigma}(0, 0) = 0$ this implies that $g_m(\lambda) = \Lambda_{W,\sigma}(\lambda, -m\lambda) < 0$ for $\lambda < 0$ sufficiently close to 0, and therefore $\limsup_{\lambda \to 0^-} \Lambda_V(\lambda)/\lambda \geq m$. Since this is true for any $m < \infty$ and since $\Lambda_V(\lambda)$ is convex, it follows that $\lim_{\lambda \to 0^-} \Lambda_V(\lambda)/\lambda = \infty$. Thus, for any $x < \infty$ there exists a $\lambda' < 0$ such that $\Lambda_V(\lambda') < \lambda' x$ and so $I_V(x) \geq \lambda' x - \Lambda_V(\lambda') > 0$.

**Case II:** $\delta \leq 1$.

As in Case I we could proceed by arguing that $g'_m(\lambda) \to E[(W_1 - m\sigma_1)\mathbb{1}_{\{\sigma_1 < \infty\}}]$. However, we would need to then show that this last expectation is infinite, and this would require an analysis of the joint tail behavior of $(W_1, \sigma_1)$. This could probably be achieved in the case $\delta \in (0, 1)$ by adapting the arguments of Kosygina and Mountford in [13], however when $\delta < 0$ it would be more difficult since in that case the Markov chain is transient and we would need to analyze the tails of $\sigma_1$ conditioned on $\sigma_1 < \infty$. It is possible that such an approach would work, but we will give a softer argument instead.

Let $\Lambda_1(\lambda) = \limsup_{n \to \infty} \frac{1}{n} \log E[e^{\lambda \sum_{i=1}^n V_i}]$. Then, the standard Chebyshev large deviation upper bound implies that

$$\limsup_{n \to \infty} \frac{1}{n} \log P \left( \sum_{i=1}^n V_i < xn \right) \leq -\sup_{\lambda < 0} (\lambda x - \Lambda_1(\lambda)).$$

On the other hand, Theorem 3.1 and the fact that $I_V$ is non-increasing implies that

$$\lim_{n \to \infty} \frac{1}{n} \log P \left( \sum_{i=1}^n V_i < xn, V_n = 0 \right) = -I_V(x).$$

Thus, we see that $I_V(x) \geq \sup_{\lambda<0}(\lambda x - \Lambda_1(\lambda))$ for any $x < \infty$. Then, similar to the case $\delta \in (1,2]$ above, it will follow that $I_V(x) > 0$ for all $x < \infty$ if we can show that $\lim_{\lambda\to 0^-} \Lambda_1(\lambda)/\lambda = \infty$.

Fix an integer $K \geq 1$. If $\lambda < 0$, then $\lambda \sum_{i=1}^n V_i \leq \lambda K \sum_{i=1}^n \mathbb{1}_{\{V_i \geq K\}}$. Thus,

$$E[e^{\lambda \sum_{i=1}^n V_i}] \leq e^{\lambda K(1-\theta)n} + P\left(\sum_{i=1}^n \mathbb{1}_{\{V_i < K\}} > \theta n\right). \tag{3.15}$$

Recall the construction of the process $V_i$ in Section 2 and define $\tilde{V}_i$ by $\tilde{V}_0 = 0$ and

$$\tilde{V}_{i+1} = F^{(i)}_{\tilde{V}_i+1} \mathbb{1}_{\{F^{(i)}_{\tilde{V}_i+1} \geq K\}}.$$

That is, jumps are governed by the same process as the jumps of the Markov chain $V_i$ with the exception that any attempted jump to a site in $[1, K-1]$ is replaced by a jump to 0. Note that the above construction of $\tilde{V}_i$ gives a natural coupling with $V_i$ so that $\tilde{V}_i \leq V_i$ for all $i$. Let $\tilde{\sigma}_k$, $k = 1, 2, \ldots$ be the successive return times to 0 of the Markov chain $\tilde{V}_i$. Then, since $\tilde{V}_i < K$ implies that $\tilde{V}_i = 0$,

$$P\left(\sum_{i=1}^n \mathbb{1}_{\{V_i < K\}} > \theta n\right) \leq P\left(\sum_{i=1}^n \mathbb{1}_{\{\tilde{V}_i < K\}} > \theta n\right) \leq P(\tilde{\sigma}_{\lceil \theta n \rceil} \leq n).$$

Since $\tilde{\sigma}_k$ is the sum of $k$ i.i.d. random variables, Cramer's Theorem implies that this last probability decays on an exponential scale like $e^{-n\theta I_{\tilde{\sigma}}(1/\theta)}$, where $I_{\tilde{\sigma}}$ is the large deviation rate function for $\tilde{\sigma}_k/k$.

Recalling (3.15), we see that $\Lambda_1(\lambda) \leq -\min\{\lambda K(\theta - 1), \theta I_{\tilde{\sigma}}(1/\theta)\}$. Optimizing over $\theta \in (0,1)$ gives

$$\Lambda_1(\lambda) \leq -\sup_{\theta \in (0,1)} \min\{\lambda K(\theta - 1), \theta I_{\tilde{\sigma}}(1/\theta)\}. \tag{3.16}$$

The modified Markov chain $\tilde{V}_i$ inherits the same recurrence/transience properties that $V_i$ has. In particular, $\tilde{V}_i$ is null-recurrent if $\delta \in [0,1]$ and transient if $\delta < 0$. In either case $E[\tilde{\sigma}_1] = \infty$ and so it can be shown that $I_{\tilde{\sigma}}(x)$ is convex, non-increasing, and $I_{\tilde{\sigma}}(x) > 0$ for $x \in [1, \infty)$. Therefore, the function $\theta \mapsto \theta I_{\tilde{\sigma}}(1/\theta)$ is convex and strictly increasing on $(0,1)$ and approaches 0 as $\theta \to 0$. Thus, there exists an inverse function $h$ so that $h(x)I_{\tilde{\sigma}}(1/h(x)) = x$ and $h(x) \to 0$ as $x \to 0$. We will use this information to analyze the upper bound in (3.16).
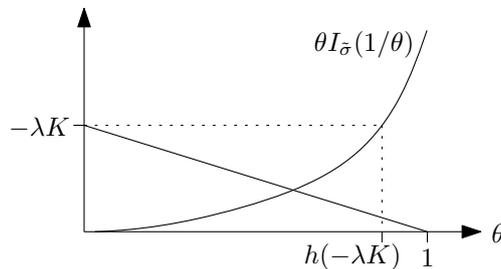


Figure 1: For any fixed $\lambda < 0$, the supremum in (3.16) is attained at the intersection of the two curves. A lower bound for the supremum is obtained by evaluating the line $\lambda K(\theta - 1)$ at $\theta = h(-\lambda K)$.

Since the first term in the minimum of (3.16) is decreasing in $\theta$ and the second term in the minimum is increasing in $\theta$, the supremum is obtained for the value of $\theta$

that makes the two terms in the minimum equal. Thus, the supremum is greater than $\lambda K(h(-\lambda K) - 1)$ (see Figure 1) which in turn implies that $\Lambda_1(\lambda) \leq \lambda K(1 - h(-\lambda K))$. Therefore,

$$\liminf_{\lambda \to 0^-} \frac{\Lambda_1(\lambda)}{\lambda} \geq \lim_{\lambda \to 0^-} K(1 - h(-\lambda K)) = K.$$

Since the above argument works for any finite $K$, this implies that $\lim_{\lambda \to 0^-} \Lambda_1(\lambda)/\lambda = \infty$. $\qquad\square$

## 4  Large Deviations for Hitting Times

The large deviation principles for $n^{-1} \sum_{i=1}^{n} V_i$ in Theorems 3.1 and 3.2 imply a large deviation principle for the hitting times.

**Theorem 4.1.** *Let $I_T(t) = I_V((t-1)/2)$. Then, $T_n/n$ satisfies a large deviation principle with convex rate function $I_T(t)$. That is,*

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left(T_n/n \in G\right) \geq - \inf_{x \in G} I_V(x), \tag{4.1}$$

*for all open $G$, and*

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left(T_n/n \in F\right) \leq - \inf_{x \in F} I_V(x), \tag{4.2}$$

*for all closed $F$. Moreover, the following qualitative properties are true of the rate function $I_T$.*

 *(i) $I_T(t)$ is convex, non-increasing, and continuous on $[1, \infty)$, and there exists a $t_2 > 1$ such that $I_T(t)$ is strictly convex and analytic on $(1, t_2)$.*

 *(ii) $I_T(1) = - \log \mathbf{E}[\omega_0(1)]$ and $\lim_{t \to 1^+} I_T'(t) = -\infty$.*

 *(iii) If $\delta > 2$, then $I_T(t) = 0 \iff t \geq 1/v_0$. Moreover, there exists a $t_1 < 1/v_0$ such that $I_T(t)$ is strictly convex and analytic on $(t_1, 1/v_0)$ and continuously differentiable on $(t_1, \infty)$.*

 *(iv) If $\delta \leq 2$, then $I_T(t) > 0$ for all $t < \infty$ and $\lim_{t \to \infty} I_T(t) = 0$.*
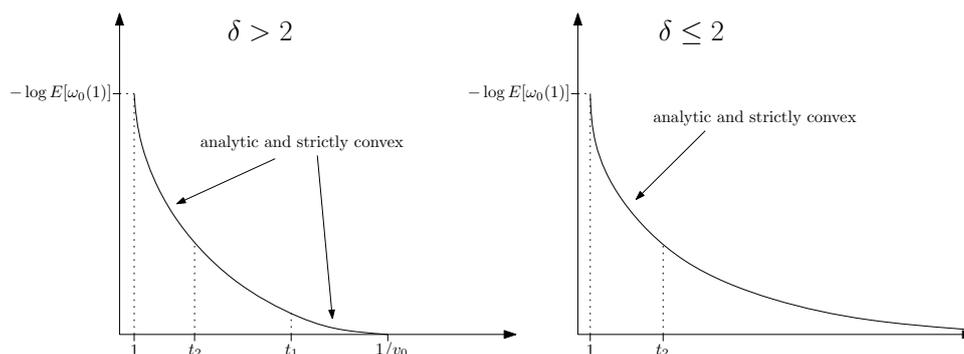


Figure 2: A visual depiction of the rate function $I_T$ in the cases $\delta > 2$ and $\delta \leq 2$ showing the qualitative properties stated in Lemma 4.1.

*Proof.* The properties of the rate function $I_T$ follow directly from the corresponding properties of $I_V$ proved above in Lemmas 3.9 and 3.10. Note that when $\delta > 2$ we use that the formula for the limiting speed of the excited random walk in (2.4) implies that $1/v_0 = E[\sigma_1 + 2W_1]/E[\sigma_1] = 1 + 2m_0$.

Recall the relationship between the hitting times $T_n$ and the processes $V_i$ and $V_i^{(n)}$ given in (2.1). Then,

$$P(T_n/n \in G) \geq P\left(1 + \frac{2}{n}\sum_{i=1}^{n} V_i \in G, \; V_n = 0\right),$$

since $V_i^{(n)} = 0$ for all $i \geq 1$ if $V_0^{(n)} = V_n = 0$. The large deviation lower bound (4.1) then follows from Theorem 3.1.

Since $I_T$ is non-increasing and $\inf_t I_T(t) = 0$, as in the proof of Theorem 3.2 the large deviation upper bound will follow from

$$\limsup_{n \to \infty} \frac{1}{n} \log P(T_n \leq nt) = -I_T(t). \tag{4.3}$$

Again, the relationship between the hitting times $T_n$ and the processes $V_i$ and $V_i^{(n)}$ given in (2.1) implies that

$$P(T_n \leq tn) \leq P\left(\sum_{i=1}^{n} V_i \leq \frac{(t-1)n}{2}\right),$$

and Theorem 3.2 implies that (4.3) holds. $\qquad\square$

To obtain a large deviation principle for the position of the excited random walk we will also need a large deviation principle for the hitting times to the left. However, this is obtained directly as a Corollary of Theorem 4.1 by switching the direction of the cookie drifts. To be more precise, for any cookie environment $\omega = \{\omega_i(j)\}$, let $\overline{\omega} = \{\overline{\omega}_i(j)\}$ be the associated cookie environment given by $\overline{\omega}_i(j) = 1 - \omega_i(j)$. Let $\overline{T}_n$ be the hitting times of the excited random walk in the cookie environment $\overline{\omega}$. An obvious symmetry coupling gives $T_{-n} = \overline{T}_n$.

**Corollary 4.2.** *The random variables $T_{-n}/n$ satisfy a large deviation principle with convex rate function $I_{\overline{T}}$, where $I_{\overline{T}}$ is the rate function given by Theorem 4.1 for the hitting times $\overline{T}_n/n$.*

**Remark 4.3.** *Since $\overline{\delta} = E[\sum_{j=1}^{M}(2\overline{\omega}_0(j) - 1)] = -E[\sum_{j=1}^{M}(2\omega_0(j) - 1)] = -\delta$, the properties of the rate function $I_{\overline{T}}$ are the same as the properties of the rate function $I_T$ given by Theorem 4.1 when $\delta$ is replaced by $-\delta$. For instance, $I_{\overline{T}}(t) > 0$ for all $t < \infty$ if $\delta \geq -2$.*

## 5  Large deviations for the random walk

In this section will show a large deviation principle for $X_n/n$. We begin by defining the rate function $I_X(x)$.

$$I_X(x) = \begin{cases} x I_T(1/x) & x > 0 \\ 0 & x = 0 \\ |x| I_{\overline{T}}(1/|x|) & x < 0. \end{cases} \tag{5.1}$$

Before stating the large deviation principle for $X_n/n$ we will prove some simple facts about the rate function $I_X$.

**Lemma 5.1.** *The function $I_X$ is non-negative and continuous on $[-1, 1]$ and has the following additional properties*

(i)  *$I_X(x)$ is non-increasing on $[-1, 0]$ and non-decreasing on $[0, 1]$.*
(ii)  *$I_X(x)$ is a convex function.*
(iii)  *$I_X(-1) = -\log \mathbf{E}[1 - \omega_0(1)]$ and $I_X(1) = -\log \mathbf{E}[\omega_0(1)]$.*

(iv) *There exist $\overline{x}_2 \in (-1,0)$ and $x_2 \in (0,1)$ such that $I_X$ is strictly convex and analytic on $(-1,\overline{x}_2)$ and $(x_2,1)$.*

(v) $\lim_{x\to -1^+} I_X'(x) = -\infty$ *and* $\lim_{x\to 1^-} I_X'(x) = \infty$.

(vi) $I_X'(0) = \lim_{x\to 0} I_X(x)/x = 0$.

(vii) *If $\delta \in [-2,2]$, then $I_X(x) = 0$ if and only if $x = 0$.*

(viii) *If $\delta > 2$, then $I_X(x) = 0$ if and only if $x \in [0, v_0]$, and there exists an $x_1 \in (v_0, 1)$ such that $I_X$ is strictly convex and analytic on $(v_0, x_1)$ and continuously differentiable on $[0, x_1)$.*

(ix) *If $\delta < -2$ then $I_X(x) = 0$ if and only if $x \in [v_0, 0]$, and there exists an $\overline{x}_1 \in (-1, v_0)$ such that $I_X$ is strictly convex and analytic on $(\overline{x}_1, v_0)$ and continuously differentiable on $(\overline{x}_1, 0]$..*
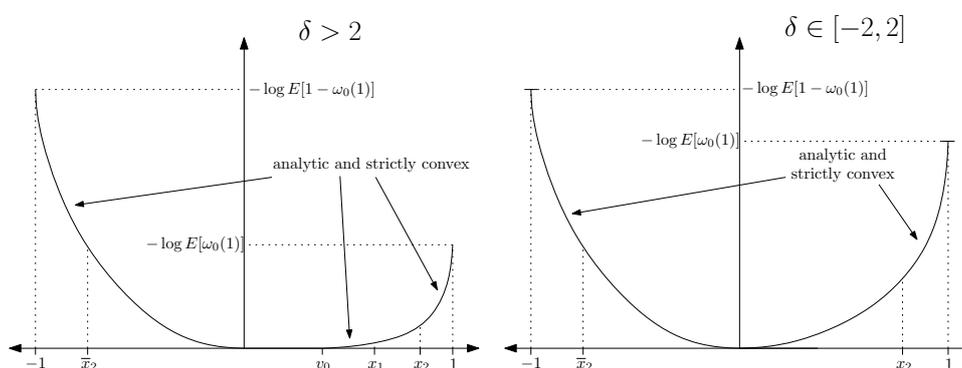


Figure 3: A visual depiction of the rate function $I_X$ in the cases $\delta > 2$ and $\delta \in [-2,2]$ showing the qualitative properties stated in Lemma 5.1.

*Proof.* Most of the properties in the statement of the Lemma follow directly from the corresponding properties of $I_T$ (or $I_{\overline{T}}$) given by Theorem 4.1, and thus we will content ourselves with only discussing property (ii) from the statement of the Lemma.

It is a general fact of convex analysis that if $f(x)$ is a convex function on $[1,\infty)$ then $g(x) = xf(1/x)$ is also a convex function on $(0,1]$. Therefore, the convexity of $I_T$ and $I_{\overline{T}}$ imply that $I_X$ is convex on $[-1,0)$ and $(0,1]$, respectively. Next, note that $\lim_{x\to 0^+} I_X(x) = \lim_{x\to 0^+} xI_T(1/x) = 0$ since $I_T$ is finite and non-increasing, and similarly $\lim_{x\to 0^-} I_X(x) = 0$. Therefore, $I_X$ is continuous at $x = 0$ which in turn implies that $I_X$ is convex on $[-1,0]$ and $[0,1]$, respectively. Finally, the convexity of $I_X$ on all of $[-1,1]$ follows from the convexity on $[-1,0]$ and $[0,1]$ and the monotonicity properties in in part (i) of the lemma. $\square$

We now are ready to prove the large deviation principle for the position of the excited random walk.

*Proof of Theorem 1.1.* Since the rate function $I_X$ is non-increasing on $[-1,0)$, $I_X(0) = 0$, and $I_X$ is non-decreasing on $(0,1]$, it is enough to prove the large deviation upper bound for closed sets of the form $F = [x,1]$ with $x > 0$ or $F = [-1,x]$ with $x < 0$. To this end, let $x > 0$ and note that $\{X_n \geq nx\} \subset \{T_{\lceil nx \rceil} \leq n\}$. Then, Theorem 4.1 implies that

$$\limsup_{n\to\infty} \frac{1}{n} \log P(X_n \geq nx) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(T_{\lceil nx \rceil} \leq n) = -xI_T(1/x), \quad \forall x \in (0,1].$$

Similarly, if $x < 0$ then $\{X_n \leq nx\} \subset \{T_{-\lceil n|x| \rceil} \leq n\}$ and Corollary 4.2 implies that

$$\limsup_{n\to\infty} \frac{1}{n} \log P(X_n \leq nx) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(T_{-\lceil n|x|\rceil} \leq n) = -|x| I_{\overline{T}}(1/|x|), \quad \forall x \in [-1, 0).$$

Recalling the definition of $I_X(x)$ in (5.1) and the monotonicity properties of $I_X$ in Lemma 5.1 finishes the proof of the large deviation upper bound.

To prove the large deviations lower bound it is enough to show that

$$\lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{n} \log P(|X_n - nx| < \varepsilon n) \geq -I_X(x), \quad \forall x \in [-1, 1], \tag{5.2}$$

First consider the case where $x \in (0, 1]$. Then, since the random walk is a nearest neighbor walk

$$P(|X_n - nx| < \varepsilon n) \geq P(|T_{\lceil nx \rceil} - n| < \varepsilon n - 1).$$

Then, Theorem 4.1 implies that for any $x \in (0, 1]$,

$$\lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{n} \log P(|X_n - nx| < \varepsilon n) \geq \lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{n} \log P(|T_{\lceil nx \rceil} - n| < \varepsilon n - 1)$$
$$\geq -x I_T(1/x) = -I_X(x),$$

and a similar argument shows that (5.2) also holds for $x \in [-1, 0)$. Finally, to show that (5.2) holds when $x = 0$ note that the naive slowdown strategy in Example 1.3 implies that $P(|X_n| \leq n^{1/3}) \geq C e^{-cn^{1/3}}$ and thus

$$\lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{n} \log P(|X_n| < \varepsilon n) = 0 = -I_X(0).$$

$\square$

## 6 Slowdowns

If $\delta > 2$, then Lemma 5.1 shows that the rate function $I_X$ is zero in the interval $[0, v_0]$. Thus, probabilities such as $P(X_n < nx)$ decay to zero sub-exponentially for $x \in (0, v_0)$. Similarly, since $I_T$ is zero in $[1/v_0, \infty)$ probabilities of the form $P(T_n > nt)$ decay sub-exponentially if $t > 1/v_0$. The main goal of this section is to prove Theorem 1.4 which gives the correct polynomial rate of decay for these probabilities.

In order to prove Theorem 1.4 we will need the following bound on backtracking probabilities for transient excited random walks.

**Lemma 6.1.** *Let $\delta > 1$. Then there exists a constant $C > 0$ such that for any $n, r \geq 1$,*

$$P\left(\inf_{k \geq T_{n+r}} X_k \leq n\right) \leq C r^{1-\delta}.$$

**Remark 6.2.** *In [2], Basdevant and Singh showed that such backtracking probabilities could be bounded uniformly in $n$ by a term that vanishes as $r \to \infty$. However, their argument uses an assumption of non-negativity of the cookie strengths, and their bounds do not give any information on the rate of decay of the probabilities in $r$. Our argument is more general (allowing positive and negative cookie drifts) and gives a quantitative rate of decay in $r$.*

*Proof.* First, note that

$$P\left(\inf_{k \geq T_{n+r}} X_k \leq n\right) = \lim_{m\to\infty} P\left(\inf_{T_{n+r} \leq k < T_m} X_k \leq n\right) \tag{6.1}$$

The event $\{\inf_{T_{n+r} \leq k < T_m} X_k \leq n\}$ implies that for every site $i \in [n+1, n+r]$ the excited random walk jumps from $i$ to $i-1$ at least one time before time $T_m$. Therefore,

$$P\left(\inf_{T_{n+r} \leq k < T_m} X_k \leq n\right) \leq P(U_i^m \geq 1, \forall i \in [n+1, n+r])$$

$$= P(V_i \geq 1, \forall i \in [m-n-r, m-n-1]). \qquad (6.2)$$

Now, the asymptotic age distribution for a discrete renewal process (see Section 6.2 of [15]) implies that for any $k \geq 1$

$$\lim_{m \to \infty} P(V_i \neq 0 \text{ for all } m < i \leq m+k) = \frac{E[(\sigma_1 - k)_+]}{E[\sigma_1]}.$$

Applying this to (6.1) and (6.2) we obtain

$$P\left(\inf_{k \geq T_{n+r}} X_k \leq n\right) \leq \frac{E[(\sigma_1 - r)_+]}{E[\sigma_1]}.$$

The tail decay of $\sigma_1$ in (2.3) implies that when $\delta > 1$ there exists a constant $C > 0$ such that $E[(\sigma_1 - r)_+] \leq Cr^{1-\delta}$ for any $r \geq 1$. $\qquad \square$

We will also need the following large deviation asymptotics for heavy tailed random variables.

**Lemma 6.3.** *Let $\{Z_k\}_{k \geq 1}$ be i.i.d. non-negative random variables with $P(Z_1 > t) \sim Ct^{-\kappa}$ for some $\kappa > 1$ and $C > 0$. Then,*

$$\lim_{n \to \infty} \frac{\log P\left(\sum_{k=1}^n Z_k > xn\right)}{\log n} = 1 - \kappa, \quad \forall x > E[Z_1].$$

**Remark 6.4.** *Lemma 6.3 is not new, but we provide a quick proof here for the convenience of the reader since we could not find a statement of this lemma in the literature.*

*Proof.* The statement of the Lemma follows easily from [17, equation (0.3)] when $\kappa > 2$. Indeed, if $\kappa > 2$ then in fact

$$P\left(\sum_{k=1}^n Z_k > xn\right) \sim nP\left(Z_1 - E[Z_1] > n(x - E[Z_1])\right)$$

$$\sim C(x - E[Z_1])^{-\kappa} n^{1-\kappa}, \quad \text{as } n \to \infty,$$

for any $x > E[Z_1]$.

When $\kappa \in (1, 2]$ we can no longer use [17, equation (0.3)] and so a different approach is needed. To this end, first note that since the $Z_k$ are non-negative a simple lower bound is

$$P\left(\sum_{k=1}^n Z_k > xn\right) \geq P(\exists k \leq n : Z_k > xn) = 1 - (1 - P(Z_1 > xn))^n.$$

Since $1 - (1-p)^n \geq np + (np)^2/2$ for any $n \geq 1$ and $p \in [0, 1]$ this implies that

$$P\left(\sum_{k=1}^n Z_k > xn\right) \geq nP(Z_1 > xn) + \frac{1}{2}n^2 P(Z_1 > xn)^2 \sim Cx^{-\kappa} n^{1-\kappa}.$$

To obtain a corresponding upper bound when $\kappa \in (1, 2]$, note that $E[Z_1^\gamma] < \infty$ for any $\gamma \in (0, \kappa)$. Then, [4] implies that $P(\sum_{k=1}^n Z_k > xn) = o(n^{1-\gamma})$ for any $\gamma \in (0, \kappa)$ and any $x > E[Z_1]$, and this is enough to complete the proof of the lemma. $\qquad \square$

We are now ready to give the proof of the main result of this section.

*Proof of Theorem 1.4.* We first prove the polynomial rate of decay for the hitting time probabilities in (1.4). Since $\sigma_k$ and $W_k$ are sums of $k$ i.i.d. non-negative random variables with tail decay given by (2.3), Lemma 6.3 implies that

$$\lim_{k\to\infty} \frac{1}{\log k} P(\sigma_k > ky) = 1 - \delta, \quad \text{if } y > E[\sigma_1], \tag{6.3}$$

and

$$\lim_{k\to\infty} \frac{1}{\log k} P(W_k > ky) = 1 - \delta/2, \quad \text{if } y > E[W_1]. \tag{6.4}$$

Recall the relationship between the hitting times $T_n$ and the branching processes $V_i$ and $V_i^{(n)}$ given in (2.1). Also, note that the branching process $V_i^{(n)}$ starts with $V_0^{(n)} = V_n$ and has the same offspring distribution as the branching process $V_i$ but without the extra immigrant each generation. Thus, $V_i^{(n)} = 0$ implies that $V_j^{(n)}$ for all $j \geq i$ and the processes are naturally coupled so that $V_i^{(n)} \leq V_{n+i}$ for all $i \geq 1$. Therefore, $T_n$ is stochastically dominated by $n + 2\sum_{i=1}^{\sigma_{k(n)}} V_i = n + 2W_{k(n)}$, where $k(n)$ is defined by $\sigma_{k(n)-1} < n \leq \sigma_{k(n)}$. Thus, for any $c > 0$

$$P(T_n > nt) \leq P(k(n) > cn) + P\left(W_{\lfloor cn \rfloor} > \frac{n(t-1)}{2}\right)$$

$$\leq P(\sigma_{\lfloor cn \rfloor} < n) + P\left(W_{\lfloor cn \rfloor} > \frac{n(t-1)}{2}\right). \tag{6.5}$$

While (6.3) implies that the right tail large deviations of $\sigma_k/k$ decay polynomially, the left tail large deviations decay exponentially since $\sigma_k$ is the sum of non-negative random variables (use Cramer's theorem). That is,

$$\lim_{k\to\infty} \frac{1}{k} \log P(\sigma_k < ky) < 0, \quad \text{if } y < E[\sigma_1].$$

Therefore, if we can choose $c$ such that $1/c < E[\sigma_1]$ and $(t-1)/(2c) > E[W_1]$ the first term in (6.5) will decay exponentially in $n$ while the second term will decay polynomially on the order $n^{1-\delta/2}$. The assumption that $t > 1/v_0 = 1 + 2E[W_1]/E[\sigma_1]$ implies that $(t-1)/2 > E[W_1]/E[\sigma_1]$ and so such a $c$ may be found.

For a matching lower bound on the polynomial rate of decay of $P(T_n > nt)$, we again use the relationship between the hitting times and the branching process in (2.1) to obtain

$$P(T_n > nt) \geq P\left(\sum_{i=1}^{n} V_i > \frac{n(t-1)}{2}\right)$$

$$\geq P\left(\exists k \leq n : W_k > \frac{n(t-1)}{2}, \, \sigma_k \leq n\right)$$

$$\geq P\left(W_{cn} > \frac{n(t-1)}{2}\right) - P(\sigma_{cn} > n).$$

If $c < (E[\sigma_1])^{-1}$ then the assumption that $t > 1/v_0$ implies that $(t-1)/(2c) > E[W_1]$, and so (6.4) and (6.3) imply that $P(T_n > nt) \geq n^{1-\delta/2+o(1)} - n^{1-\delta+o(1)} = n^{1-\delta/2+o(1)}$. This completes the proof of (1.4).

We now turn to the subexponential rate of decay for $P(X_n < xn)$. A lower bound follows immediately from (1.4) since $P(X_n < xn) \geq P(T_{\lceil xn \rceil} > n)$. To obtain a corre-

sponding upper bound, note that

$$P(X_n < xn) \le P(T_{\lceil n(x+\varepsilon)\rceil} > n) + P\left(\inf_{k > T_{\lceil n(x+\varepsilon)\rceil}} X_k < xn\right)$$

$$\le P(T_{\lceil n(x+\varepsilon)\rceil} > n) + C(n\varepsilon)^{1-\delta}, \tag{6.6}$$

where the last inequality follows from Lemma 6.1. Now, if $\varepsilon > 0$ is sufficiently small (so that $x + \varepsilon < v_0$) then (1.4) implies that the probability in (6.6) is $n^{1-\delta/2+o(1)}$. Since $n^{1-\delta/2}$ is much larger than $n^{1-\delta}$ this completes the proof of the upper bound needed for (1.3). $\square$

# References

[1] Anne-Laure Basdevant and Arvind Singh, *On the speed of a cookie random walk*, Probab. Theory Related Fields **141** (2008), no. 3-4, 625–645. MR-2391167

[2] _____, *Rate of growth of a transient cookie random walk*, Electron. J. Probab. **13** (2008), no. 26, 811–851. MR-2399297

[3] Itai Benjamini and David B. Wilson, *Excited random walk*, Electron. Comm. Probab. **8** (2003), 86–92 (electronic). MR-1987097

[4] David R. Brillinger, *A note on the rate of convergence of a mean*, Biometrika **49** (1962), 574–576. MR-0156372

[5] Wlodzimierz Bryc and Amir Dembo, *Large deviations and strong mixing*, Ann. Inst. H. Poincaré Probab. Statist. **32** (1996), no. 4, 549–569. MR-1411271

[6] Francis Comets, Nina Gantert, and Ofer Zeitouni, *Quenched, annealed and functional large deviations for one-dimensional random walk in random environment*, Probab. Theory Related Fields **118** (2000), no. 1, 65–114. MR-1785454

[7] A. de Acosta, *Upper bounds for large deviations of dependent random vectors*, Z. Wahrsch. Verw. Gebiete **69** (1985), no. 4, 551–565. MR-791911

[8] Amir Dembo, Yuval Peres, and Ofer Zeitouni, *Tail estimates for one-dimensional random walk in random environment*, Comm. Math. Phys. **181** (1996), no. 3, 667–683. MR-1414305

[9] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition. MR-2571413

[10] Dmitry Dolgopyat, *Central limit theorem for excited random walk in the recurrent regime*, ALEA Lat. Am. J. Probab. Math. Stat. **8** (2011), 259–268. MR-2831235

[11] Dmitry Dolgopyat and Elena Kosygina, *Scaling limits of recurrent excited random walks on integers*, January 2012, available at arXiv:math/1201.0379.

[12] H. Kesten, M. V. Kozlov, and F. Spitzer, *A limit law for random walk in a random environment*, Compositio Math. **30** (1975), 145–168. MR-0380998

[13] Elena Kosygina and Thomas Mountford, *Limit laws of transient excited random walks on integers*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 2, 575–600. MR-2814424

[14] Elena Kosygina and Martin P. W. Zerner, *Positively and negatively excited random walks on integers, with branching processes*, Electron. J. Probab. **13** (2008), no. 64, 1952–1979. MR-2453552

[15] Gregory F. Lawler, *Introduction to stochastic processes*, second ed., Chapman & Hall/CRC, Boca Raton, FL, 2006. MR-2255511

[16] A. A. Mogul′skiĭ, *Small deviations in the space of trajectories*, Teor. Verojatnost. i Primenen. **19** (1974), 755–765. MR-0370701

[17] S. V. Nagaev, *Large deviations of sums of independent random variables*, Ann. Probab. **7** (1979), no. 5, 745–789. MR-542129

[18] P. Ney and E. Nummelin, *Markov additive processes II. Large deviations*, Ann. Probab. **15** (1987), no. 2, 593–609. MR-885132

[19] Firas Rassoul-Agha, *Large deviations for random walks in a mixing random environment and other (non-Markov) random walks*, Comm. Pure Appl. Math. **57** (2004), no. 9, 1178–1196. MR-2059678

[20] Fred Solomon, *Random walks in a random environment*, Ann. Probability **3** (1975), 1–31. MR-0362503

[21] Martin P. W. Zerner, *Multi-excited random walks on integers*, Probab. Theory Related Fields **133** (2005), no. 1, 98–122. MR-2197139