Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2017, Vol. 53, No. 3, 1458–1497
DOI: 10.1214/16-AIHP761

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Excited random walks with Markovian cookie stacks

Elena Kosygina[a,1] and Jonathon Peterson[b,2]

[a] *Department of Mathematics, Baruch College, One Bernard Baruch Way, Box B6-230, New York, NY 10010, USA.*
E-mail: elena.kosygina@baruch.cuny.edu; url: www.baruch.cuny.edu/math/elenak/
[b] *Department of Mathematics, Purdue University, 150 N University Street, Lafayette, IN 47907, USA.*
E-mail: peterson@purdue.edu; url: www.math.purdue.edu/~peterson

Received 23 April 2015; revised 26 March 2016; accepted 30 April 2016

Abstract. We consider a nearest-neighbor random walk on $\mathbb{Z}$ whose probability $\omega_x(j)$ to jump to the right from site x depends not only on x but also on the number of prior visits j to x. The collection $(\omega_x(j))_{x\in\mathbb{Z},j\geq 1}$ is sometimes called the "cookie environment" due to the following informal interpretation. Upon each visit to a site the walker eats a cookie from the cookie stack at that site and chooses the transition probabilities according to the "strength" of the cookie eaten. We assume that the cookie stacks are i.i.d. and that the cookie "strengths" within the stack $(\omega_x(j))_{j\geq 1}$ at site x follow a finite state Markov chain. Thus, the environment at each site is dynamic, but it evolves according to the local time of the walk at each site rather than the original random walk time.

The model admits two different regimes, critical or non-critical, depending on whether the expected probability to jump to the right (or left) under the invariant measure for the Markov chain is equal to $1/2$ or not. We show that in the non-critical regime the walk is always transient, has non-zero linear speed, and satisfies the classical central limit theorem. The critical regime allows for a much more diverse behavior. We give necessary and sufficient conditions for recurrence/transience and ballisticity of the walk in the critical regime as well as a complete characterization of limit laws under the averaged measure in the transient case.

The setting considered in this paper generalizes the previously studied model with periodic cookie stacks [Excited random walk with periodic cookies (2014) Preprint]. Our results on ballisticity and limit theorems are new even for the periodic model.

Résumé. Nous considérons une marche aléatoire au plus proche voisin sur $\mathbb{Z}$ dont la probabilité $\omega_x(j)$ de sauter à droite du site x ne dépend pas seulement de x mais aussi du nombre j de visites antérieures en x. La collection $(\omega_x(j))_{x\in\mathbb{Z},j\geq 1}$ est parfois nommée « l'environnement cookie » à cause de l'interprétation suivante. À chaque visite d'un site le marcheur mange un cookie de la pile de cookie à ce site et choisi la probabilitéé de transition en fonction de la force du cookie qui a été mangé. Nous supposons que les piles de cookie sont i.i.d. et que la force des cookies à l'intérieur de la pile $(\omega_x(j))_{j\geq 1}$ au site x est une chaine de Markov à espace d'états fini. Par conséquent l'environnement à chaque site est dynamique mais évolue en fonction du temps local de la marche à chaque site, plutôt que le temps propre de la marche aléatoire originale.

Le modèle admet deux régimes différents, critique ou non critique, dépendant du fait que la probabilité sous la mesure invariante de la chaine de Markov de sauter à droite (ou à gauche) est égale à $1/2$ ou non. Nous montrons que dans le régime non-critique la marche est toujours transiente, a une vitesse déchappement linéaire et satisfait le théorème de la limite centrale. Le régime critique a beaucoup plus de variantes possibles. Nous donnons alors des conditions nécessaires et suffisantes pour la recurrence/transience de la marche et une caractérisation complète des lois limites possibles sous la mesure moyennisée dans le cas transient.

Le cadre de ce papier généralise le modèle étudié précédemment où les piles de cookies étaient périodiques [Excited random walk with periodic cookies (2014) Preprint]. Nos résultats sur la ballisticité et les théorèmes limites sont nouveaux même pour le modèle périodique.

MSC: Primary 60K37; secondary 60F05; 60J10; 60J15; 60K35

[1] Supported in part by the Simons Foundation through a Collaboration Grant for Mathematicians #209493 and Simons Fellowship in Mathematics, 2014–2015.
[2] Supported in part by NSA Grant H98230-13-1-0266.

Keywords: Excited random walk; Diffusion approximation; Stable limit laws; Random environment; Branching-like processes

1. Introduction

Excited random walks (ERWs, also called cookie random walks) are a model for a self-interacting random motion where the self-interaction is such that the transition probabilities for the next step of the random walk depend on the local time at the current location. To make this precise, a *cookie environment* $\omega = \{\omega_x(j)\}_{x\in\mathbb{Z}, j\geq 1}$ is an element of $\Omega = [0,1]^{\mathbb{Z}\times\mathbb{N}}$, and given a cookie environment $\omega \in \Omega$ and $x \in \mathbb{Z}$, the ERW in the cookie environment ω started at x is a stochastic process $\{X_n\}_{n\geq 0}$ with law P_ω^x such that $P_\omega^x(X_0 = x) = 1$ and

$$\begin{aligned} P_\omega^x(X_{n+1} = X_n + 1 | X_0, X_1, \dots, X_n) &= 1 - P_\omega^x(X_{n+1} = X_n - 1 | X_0, X_1, \dots, X_n) \\ &= \omega_{X_n}\big(\#\{k \leq n : X_k = X_n\}\big). \end{aligned}$$

The "cookie" terminology in the above description of ERWs comes from the following interpretation. One imagines a stack of cookies at each site $x \in \mathbb{Z}$. Upon the jth visit to the site x the random walker eats the jth cookie at that site which induces a step to the right with probability $\omega_x(j)$ and to the left with probability $1 - \omega_x(j)$. For this reason we will refer to $\omega_x(j)$ as the *strength* of the jth cookie at site x.

Many results for ERWs in one dimension were obtained under the assumption that there is an $M < \infty$ such that $\omega_x(j) = 1/2$ for all $j > M$ and $x \in \mathbb{Z}$ [2–5,8,23,25,30–33]. We will refer to this as the case of "boundedly many cookies per site." Much less is known if the number of cookies per site is not bounded (in particular, if there are infinitely many cookies at each site). Notable exceptions are that Zerner [42] proved a criterion for recurrence/transience and Dolgopyat [7] proved the scaling limits for the recurrent case under the assumption that $\omega_x(j) \geq 1/2$ for all $x \in \mathbb{Z}$ and $j \geq 1$ (i.e., all cookies have non-negative drift). For a review (prior to 2012) of ERWs in one and more dimensions see [24].

Until recently, very little was known for ERWs which had infinitely many cookies per site with both positive and negative drift cookies.[3] However, [27] gave an explicit criterion for recurrence/transience of ERWs in cookie environments such that the cookie sequence $\{\omega_x(j)\}_{j\geq 1}$ is the same for all x and is periodic in j. The results in the current paper cover more general cookie environments where the cookie stack at each site is given by a finite state Markov chain. Moreover, we add to the results of [27] by also proving a criterion for ballisticity (non-zero limiting velocity) and limiting distributions in the transient case. We also note that our model is general enough to include the case of periodic cookie stacks as in [27] as well as some cases of boundedly many cookies per site.

1.1. *Description of the model*

Let $\mathcal{R} = \{1, 2, \dots, N\}$ denote a finite state space, and let $\{R_j\}_{j\geq 1}$ be a Markov chain on $\mathcal{R}$ with transition probabilities given by an $N \times N$ matrix $K = (K_{i,i'})_{i,i'\in\mathcal{R}}$. We will assume that this Markov chain has a unique closed irreducible subset $\mathcal{R}_0 \subseteq \mathcal{R}$. Thus, there is a unique stationary distribution μ for the Markov chain $\{R_j\}_{j\geq 1}$, and μ is supported on $\mathcal{R}_0$. Throughout the paper we will often represent probability distributions η on $\mathcal{R}$ as row vectors $\eta = (\eta(1), \eta(2), \dots, \eta(N))$, so that for the stationary distribution μ we have $\mu K = \mu$.

Assumptions (i.i.d., elliptic, Markovian cookie stacks). *Let $\{\mathbf{R}^x\}_{x\in\mathbb{Z}}$ be an i.i.d. family of Markov chains $\mathbf{R}^x = \{R_j^x\}_{j\geq 1}$ with transition matrix K. We assume that the cookie environment ω is given by $\omega_x(j) = p(R_j^x)$ for some fixed function $p : \mathcal{R} \to (0,1)$. The requirement that p is strictly between* 0 *and* 1 *will be referred to as ellipticity throughout the paper.*

For any distribution η on $\mathcal{R}$, let $\mathbb{P}_\eta$ denote the distribution of $\{R_j^x\}_{x\in\mathbb{Z}, j\geq 1}$ when each R_1^x, $x \in \mathbb{Z}$, has distribution η. With a slight abuse of notation $\mathbb{P}_\eta$ will also be used for the induced distribution on environments ω which is constructed

[3] With the exception of one-dimensional random walk in random environment which can be seen as a particular case of ERW where $\omega_x(j) = \omega_x(1)$ for all $j \geq 1$ at each $x \in \mathbb{Z}$.

as in the assumptions above. If η is concentrated on a single state $i \in \mathcal{R}$, i.e. if $\eta = \delta_i$, we shall use $\mathbb{P}_i$ instead of $\mathbb{P}_{\delta_i}$. Expectations with respect to $\mathbb{P}_\eta$ or $\mathbb{P}_i$ will be denoted by $\mathbb{E}_\eta$ and $\mathbb{E}_i$, respectively.

As stated above, the law of the ERW in a fixed cookie environment ω is denoted P_ω^x. We will refer to this as the *quenched* law. If the cookie environment has distribution $\mathbb{P}_\eta$, then we can also define the *averaged* law of the ERW by $P_\eta^x(\cdot) = \mathbb{E}_\eta[P_\omega^x(\cdot)]$. Expectations with respect to the quenched and averaged measures will be denoted by E_ω^x and E_η^x, respectively. Again, if η is concentrated on $i \in \mathcal{R}$ we shall set $P_i^x(\cdot) := \mathbb{E}_i[P_\omega^x(\cdot)]$. We will often be interested only in ERWs started at $X_0 = 0$, and so we will use the notation P_ω, P_η, or P_i in place of P_ω^0, P_η^0 or P_i^0, respectively.

Example 1.1 (Periodic cookie sequences). *The model above clearly generalizes the case of periodic cookie sequences at each site as in* [27]. *To obtain periodic cookie sequences set $K_{i,i+1} = 1$ for $i = 1, \dots, N-1$, $K_{N,1} = 1$, and let $\eta = \delta_1$.*

Example 1.2 (Cookies stacks of geometric height). *Fix $\alpha \in (0,1]$, $p(1) > 1/2$, and $p(2) = 1/2$. Let $K = \begin{pmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{pmatrix}$. If $\eta = \delta_1$ then the cookie environment is such that there are an i.i.d.* Geometric(α) *number of cookies of strength $p(1) > 1/2$ at each site.*

Example 1.3 (Bounded cookie stacks). *The model above also generalizes the case of finitely many cookies per site. For instance, let the transition matrix K be such that $K_{i,i+1} = 1$ for $i = 1, \dots, N-1$ and $K_{N,N} = 1$, and let the function $p : \mathcal{R} \to (0,1)$ be such that $p(N) = 1/2$. In this case, if $\eta = \delta_1$ then $\omega_x(j) = p(j)$ if $j \le N-1$ and $p(j) = 1/2$ if $j \ge N$. With a little more thought, one can also obtain random cookie environments that are i.i.d. spatially with a bounded number of cookies at each site as in* [25] *subject to the restrictions that there are only finitely many possible values for the cookie strengths $\omega_x(j)$ and that $\omega_x(j) \in (0,1)$. For instance, if*

$$K = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{p} = \begin{pmatrix} p(1) \\ p(2) \\ p(3) \\ p(4) \\ 1/2 \end{pmatrix}, \quad \textit{and} \quad \eta = (\alpha, 0, 1-\alpha, 0, 0),$$

then the corresponding cookie environments under the distribution $\mathbb{P}_\eta$ have 2 *cookies at each site, having stack $(p(1), p(2))$ with probability α and stack $(p(3), p(4))$ with probability* $1-\alpha$.

Before stating our main results, we note two basic facts for ERWs that are known to hold in much more generality than our model. We will use these facts as needed throughout the paper.

Theorem 1.4 ([1, Theorem 1.2] and [24, Theorem 4.1]).

(i) *Zero-one law for transience. For any initial distribution η on $\mathcal{R}$*

$$P_\eta\Big(\lim_{n\to\infty} X_n = \infty\Big), \qquad P_\eta\Big(\lim_{n\to\infty} X_n = -\infty\Big) \in \{0,1\}.$$

(ii) *Strong law of large numbers. For any initial distribution η on $\mathcal{R}$ there is a deterministic* $v_0 \in [-1,1]$ *such that*

$$P_\eta\Big(\lim_{n\to\infty} X_n/n = v_0\Big) = 1.$$

1.2. *Main results*

The function $p : \mathcal{R} \to (0,1)$ used in the construction of the cookie environment ω corresponds to a column vector $\mathbf{p} = (p(1), p(2), \dots, p(N))^t$ with ith entry given by $p(i)$. Note that our assumptions on the Markov chains $\{\mathbf{R}^x\}_{x\in\mathbb{Z}}$, imply that the limiting average cookie strength at each site is

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \omega_x(j) = \mu \cdot \mathbf{p} =: \bar{p},$$

and that this limit does not depend on the distribution η of $\omega_x(1)$. It is natural to suspect that the ERW is transient to the right (resp. left) when $\bar{p} > 1/2$ (resp. $\bar{p} < 1/2$). Our first main result below confirms this is the case. Moreover, we also establish that if $\bar{p} \neq 1/2$ the walk has non-zero linear speed and a Gaussian limiting distribution.

Theorem 1.5. *Assume that* $\bar{p} \neq 1/2$.

Transience: *If* $\bar{p} > 1/2$ *then* $P_\eta(\lim_{n\to\infty} X_n = \infty) = 1$ *for any initial distribution* η *on* $\mathcal{R}$. *Similarly, if* $\bar{p} < 1/2$ *then* $P_\eta(\lim_{n\to\infty} X_n = -\infty) = 1$ *for any* η.

Ballisticity: *For any initial distribution* η *on* $\mathcal{R}$ *the limiting velocity* v_0 (*see Theorem* 1.4(ii)) *is non-zero*: *it is positive for* $\bar{p} > 1/2$ *and negative for* $\bar{p} < 1/2$.

Gaussian limit: *For any distribution* η *on* $\mathcal{R}$ *there exists a constant* $b = b(K, \mathbf{p}, \eta) > 0$ *such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{X_n - n\mathrm{v}_0}{b\sqrt{n}} \le x\right) = \Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz.$$

The case $\bar{p} = 1/2$ is more interesting (we will refer to this as the critical case). In fact, it will follow from our results below that ERWs in the critical case $\bar{p} = 1/2$ can exhibit the full range of behaviors that are known for ERWs with boundedly many cookies per site (e.g., transience with sublinear speed and non-Gaussian limiting distributions). Before stating our results for the critical case, we remark that in the case of boundedly many cookies per site the classification of the long term behavior of the ERW is based on a single parameter $\delta = \mathbb{E}_\eta[\sum_{j\ge 1}(2\omega_x(j) - 1)]$ which is the expected total drift contained in a cookie stack. Our results below depend on two parameters δ and $\tilde{\delta}$ which are also explicit but do not admit such simple interpretation. However, if we denote by $\mathcal{E}_0^n$ the number of upcrossings (steps to the right) the ERW makes from 0 before the nth downcrossing (step to the left) from 0 then

$$\delta = \delta(\eta, K, \mathbf{p}) := \frac{2\rho}{\nu}, \quad \text{where } \rho := \lim_{n\to\infty}\left(\mathbb{E}_\eta\left[\mathcal{E}_0^n\right] - n\right), \quad \text{and} \quad \nu := \lim_{n\to\infty} \frac{\operatorname{Var}_\eta(\mathcal{E}_0^n)}{n}.$$

The parameter $\tilde{\delta}$ is defined in a symmetric way using downcrossings instead of upcrossings and vice versa so that $\tilde{\delta} := \delta(\eta, K, \mathbf{1} - \mathbf{p})$. The explicit formulas for the parameters are not intuitive and are given in (37) and (38). In the special case of boundedly many cookies per site the parameter δ agrees with the one used in previous papers, $\tilde{\delta} = -\delta$, and $\nu = 2$. Proposition 4.3 below shows that in general $\delta + \tilde{\delta} = 1 - 2/\nu$ where ν need not be equal to 2 (see Example 1.13), and so two parameters are needed in the general model.

Theorem 1.6. *Fix an arbitrary initial distribution* η *on* $\mathcal{R}$ *and let* $\bar{p} = \mu \cdot \mathbf{p} = 1/2$. *Then, there exist constants* δ *and* $\tilde{\delta}$ *satisfying* $\delta + \tilde{\delta} < 1$ (*see* (37) *and* (38) *for the explicit formulas in terms of* η, K, *and* $\mathbf{p}$) *and such that*:

- *if* $\delta > 1$, *then* $P_\eta(\lim_{n\to\infty} X_n = \infty) = 1$;
- *if* $\tilde{\delta} > 1$, *then* $P_\eta(\lim_{n\to\infty} X_n = -\infty) = 1$;
- *if* $\delta \le 1$ *and* $\tilde{\delta} \le 1$ *then* $P_\eta(\liminf_{n\to\infty} X_n = -\infty, \limsup_{n\to\infty} X_n = \infty) = 1$.

The next theorem characterizes exactly when the walk is ballistic (i.e., has non-zero limiting velocity).

Theorem 1.7. *Fix an arbitrary initial distribution* η *on* $\mathcal{R}$. *Let* $\bar{p} = 1/2$, δ *and* $\tilde{\delta}$ *be as in Theorem* 1.6, *and* v_0 *be the limiting velocity* (*see Theorem* 1.4(ii)). *Then*:

- $\mathrm{v}_0 > 0$ *if and only if* $\delta > 2$.
- $\mathrm{v}_0 < 0$ *if and only if* $\tilde{\delta} > 2$.

Our final main result concerns the limiting distributions in the transient cases. We will state the results below for walks that are transient to the right ($\delta > 1$), though obvious symmetry considerations give similar limiting distributions for walks that are transient to the left ($\tilde{\delta} > 1$) by replacing δ with $\tilde{\delta}$. The theorem below not only gives limiting distributions for the position X_n of the ERW, but also for the hitting times T_n, where for any $x \in \mathbb{Z}$ the hitting time

$T_x = \inf\{k \geq 0 : X_k = x\}$. For $\alpha \in (0,2)$ and $b > 0$ we will use $L_{\alpha,b}(\cdot)$ to denote the distribution of the totally-skewed to the right α-stable distribution with characteristic exponent

$$\log \int_{\mathbb{R}} e^{iux} L_{\alpha,b}(dx) = \begin{cases} -b|u|^\alpha (1 - i \tan(\frac{\pi\alpha}{2}) \operatorname{sign}(u)), & \alpha \neq 1, \\ -b|u|(1 + \frac{2i}{\pi} \log|u| \operatorname{sign}(u)), & \alpha = 1. \end{cases} \tag{1}$$

Also, as in Theorem 1.5 we will use $\Phi(\cdot)$ to denote the distribution function of a standard normal random variable.

Theorem 1.8. *Let $\bar{p} = 1/2$ and suppose that $\delta > 1$.*

(i) *If $\delta \in (1,2)$, then there exists a constant $b > 0$ such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{T_n}{n^{2/\delta}} \leq x\right) = L_{\delta/2,b}(x) \quad \text{and} \quad \lim_{n\to\infty} P_\eta\left(\frac{X_n}{n^{\delta/2}} \leq x\right) = 1 - L_{\delta/2,b}\big(x^{-2/\delta}\big).$$

(ii) *If $\delta = 2$, then there exist constants $a, b > 0$ and sequences $D(n) \sim a^{-1}\log n$ and $\Gamma(n) \sim an/\log(n)$ such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{T_n - nD(n)}{n} \leq x\right) = L_{1,b}(x) \quad \text{and} \quad \lim_{n\to\infty} P_\eta\left(\frac{X_n - \Gamma(n)}{n/(\log n)^2} \leq x\right) = 1 - L_{1,b}\big(-x/a^2\big).$$

(iii) *If $\delta \in (2,4)$, then there exists a constant $b > 0$ such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{T_n - n/v_0}{n^{2/\delta}} \leq x\right) = L_{\delta/2,b}(x) \quad \text{and} \quad \lim_{n\to\infty} P_\eta\left(\frac{X_n - nv_0}{v_0^{1+2/\delta} n^{2/\delta}} \leq x\right) = 1 - L_{\delta/2,b}(-x).$$

(iv) *If $\delta = 4$, then there exists a constant $b > 0$ such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{T_n - n/v_0}{b\sqrt{n\log n}} \leq x\right) = \Phi(x) \quad \text{and} \quad \lim_{n\to\infty} P_\eta\left(\frac{X_n - nv_0}{bv_0^{3/2}\sqrt{n\log n}} \leq x\right) = \Phi(x).$$

(v) *If $\delta > 4$, then there exists a constant $b > 0$ such that*

$$\lim_{n\to\infty} P_\eta\left(\frac{T_n - n/v_0}{b\sqrt{n}} \leq x\right) = \Phi(x) \quad \text{and} \quad \lim_{n\to\infty} P_\eta\left(\frac{X_n - nv_0}{bv_0^{3/2}\sqrt{n}} \leq x\right) = \Phi(x).$$

1.3. *An overview of ideas and open questions*

Our approach is based on the following three ingredients which appeared in the literature 65–20 years ago and were successfully used by other authors in a number of different contexts: (1) mappings between random walk paths and branching process trees, (2) diffusion approximations of branching-like processes (BLPs) and Ray–Knight theorems, and (3) embedded renewal structures for BLPs.

(1) *Mappings between random walk paths and branching process trees.* The existence of a bijection between excursions of nearest-neighbor paths on $\mathbb{Z}$ and rooted trees is well known and was observed as early as [16, Section 6]. In this bijection, if the rooted tree is viewed as a genealogical tree then the offspring of the individuals in the nth generation corresponds to the number of times the nearest-neighbor path on $\mathbb{Z}$ steps from n to $n+1$ in between steps from n to $n-1$. Properties of the random walk can then be deduced from properties of the corresponding random trees. For instance, the maximum distance of the excursion from 0 corresponds to the lifetime of the branching process, and the return time of the walk to the origin is equal to twice the total progeny of the branching process over its lifetime.

Similarly, a related bijection is known to exist between rooted trees and nearest-neighbor paths in $\mathbb{Z}$ from 0 to n. In this bijection, the rooted tree corresponds to a genealogical tree where there is one immigrant per generation (for the first n generations only) and the offspring of individuals in the ith generation correspond to the number of steps from $n-i$ to $n-i-1$ between steps from $n-i$ to $n-i+1$. As with the first bijection, this bijection also can be used to deduce properties of a random walk from properties of the corresponding random tree. For instance, the hitting time

$T_n = \inf\{k \geq 1 : X_k = n\}$ is equal to n (the total number of immigrants) plus twice the total number of progeny of the branching process over its lifetime.

For either of the above bijections, putting a probability measure on random walk paths induces a probability measure on trees and vice versa. For instance, for simple symmetric random walks on $\mathbb{Z}$, the first bijection gives a measure on trees corresponding to a critical Galton–Watson process with Geometric$(1/2)$ offspring distribution, and the second bijection corresponds to a Galton–Watson process with Geometric$(1/2)$ offspring distribution but with an additional immigrant in the first n generations. These bijections were at the core of F. Knight's proof of the classical Ray–Knight theorem in [22]. For one-dimensional random walks in a random environment (RWRE), the corresponding measures on trees are instead branching processes with random offspring distributions and the second bijection above was used in [21] to obtain limiting distributions for transient one-dimensional RWRE (under the averaged measure). The advantage of using this bijection to study the RWRE is that while the random walk is non-Markovian under the averaged measure, the corresponding branching process with random offspring distribution is a Markov chain. Subsequently, it has also been observed that several other models of self-interacting random walks also have this Markovian property for the BLPs which come from the above bijections. In particular, this applies to a large number of self-interacting random walk models where the transition probabilities for the walk depend on the past behavior of the walk at the present site. Included in this are several models of self-repelling/attracting random walks [35,36,39,40] as well as excited random walks with boundedly many cookies per site. For ERWs, this branching process approach was first used in [3] and has since been the basis for many of the subsequent results on ERWs; see for instance the review [24] as well as more recent works in [2,26,27,33].

(2) *Diffusion approximations of BLPs and Ray–Knight theorems.* Diffusion approximations of branching processes is an extensively studied topic, especially in the context of applications to population dynamics (see, for example, books [19,28]). The fact that a rescaled critical Galton–Watson process converges weakly to a diffusion was first observed in [13] and then rigorously proven in [20]. In particular, a Galton–Watson process with Geometric$(1/2)$ offspring distribution converges to one-half of a zero-dimensional squared Bessel process. Adding an immigrant in each generation raises the above dimension to two so that the limiting process becomes one-half of the square of a standard two-dimensional Brownian motion. Recalling the connection of simple symmetric random walks with critical Galton–Watson processes from the bijections given above and the fact that Brownian motion is the scaling limit of random walks, we can think about the first and second classical Ray–Knight theorems as continuous versions of these bijections.[4]

For a class of non-Markovian self-interacting random motions, a generalized Ray–Knight theory was developed by Bálint Tóth (see [41] and references therein). For these walks the corresponding diffusion process limits for the local times are squared Bessel processes of appropriate dimensions. Using this generalized Ray–Knight theory the author obtains limiting distributions for the random walk stopped at an independent geometric time. Additionally, for a certain sub-class of these self-interacting random walks he identifies these limiting distributions as those of a Brownian motion perturbed at extrema [40, Remark on p. 1334] and asks if the observed connection extends to multi-dimensional distributions.

Even though the dynamics of ERWs are quite different from those of the self-interacting random walks in [41], the corresponding rescaled BLPs also converge to squared Bessel processes (see [9,23,26], and Lemmas 6.1 and 7.1 below). Similarly to [40] these diffusion limits can be used to infer certain properties of the scaling behavior of the ERW, though alone they are not quite sufficient to obtain limiting distributions for the ERWs.

(3) *Embedded renewal structure.* To obtain limit theorems for transient ERWs we essentially follow an outline which was first presented in [21, pp. 148–150] in the context of transient one-dimensional random walks in random environments. Starting with [4], this strategy has been used in essentially all papers concerned with limit laws for transient one-dimensional ERW. The idea is to first use the bijection above which relates the hitting time T_n to n plus twice the total progeny of a BLP, and then to compare the total progeny of this BLP to a sum of i.i.d. random variables using regeneration times of the BLP. To obtain limiting distributions for the hitting times (and then also the position) of the ERW using this approach, the key is to obtain precise tail asymptotics of both the regeneration times and the total progeny between regeneration times of the BLPs.

[4]The second Ray–Knight theorem was extended from Brownian motion to a large class of symmetric (and some non-symmetric) Markov processes ([11]). These extensions found many new applications. The interested reader is referred to books [29,38] and references therein.

In [4] and [25], the necessary tail asymptotics for the BLPs were obtained using generating functions or modifications of the process which allowed for the application of known results from the literature for branching processes with migration. The approach based on squared Bessel diffusion limits of the BLPs and calculations in the spirit of gambler's ruin was proposed in [23] and developed in subsequent papers [9,26]. This approach not only eliminates the need to quote results from the branching process literature but also allows one to obtain new results about BLPs. In the current work we push the method further by lifting the limitation on the number of cookies per site at the cost of requiring a Markovian structure within the cookie stacks. We provide the required tail asymptotics in the full critical regime (Theorems 2.6 and 2.7), show the one-dimensional limit laws in the transient case, but leave the recurrent case and functional limit theorems for future work.

1.3.1. *Comments and open questions*

(i) Using the formulas in (37) and (38) below, one can explicitly calculate the parameters δ and $\tilde{\delta}$ in terms of η, K, and $\mathbf{p}$. Therefore, recurrence/transience, ballisticity, and the type of the limiting distribution can be determined for any given example. However, there is no explicit formula for the limiting velocity v_0 when $v_0 \neq 0$ or for the scaling parameters $a, b > 0$ that appear in Theorem 1.8.

Question 1.9. *What can be said about monotonicity and strict monotonicity of v_0 with respect to the cookie environment?*[5]

(ii) Theorems 1.7 and 1.8 give the first such results for ERWs with an unbounded number of cookies per site (with the exception of the special case of random walks in random environments). As mentioned above, the recurrence/transience results in Theorem 1.6 were known for excited random walks with unbounded number of cookies per site only in the special cases of non-negative cookie drifts ($\omega_x(j) \geq 1/2$ for all j) [42] or periodic cookie stacks [27]. In [27] the authors also use a BLP, but their proof of the criterion for recurrence/transience differs from ours and is based on the construction of appropriate Lyapunov functions rather than on the analysis of the extinction times of the BLP. It is possible that their proof may be extended to include our model, but such an extension is not automatic since it requires an additional strong concentration estimate (see [27, Theorem 1.3]) while our approach seems much less demanding.

(iii) Functional limit theorems have been obtained for ERWs with bounded cookie stacks under the i.i.d. and (weak) ellipticity assumptions. The transient case was handled in [25, Theorem 3] and [24, Theorems 6.6 and 6.7]. Scaling limits for the recurrent case were obtained in [7] and [8]. Excursions from the origin and occupation times of the left and right semi-axes for ERW with bounded number of cookies per site have also been studied [9,26]. We believe that similar results hold for the model considered in the current paper and leave this study for future work.

Question 1.10. *State and prove functional limit theorems for the critical transient case* ($\bar{p} = 1/2$, $\max\{\delta, \tilde{\delta}\} > 1$).

Question 1.11. *Study scaling limits and occupation times of the right and left semi-axes for the recurrent case* ($\bar{p} = 1/2$, $\max\{\delta, \tilde{\delta}\} \leq 1$).

Question 1.12. *Show that the critical ERW* (*i.e.* $\bar{p} = 1/2$) *is strongly transient*[6] *under P_η if and only if* $\max\{\delta, \tilde{\delta}\} > 3$. *Show that the non-critical ERW* (*i.e.* $p \neq 1/2$) *is always strongly transient.*

1.4. *Examples*

In this subsection we give some examples where the parameters δ and $\tilde{\delta}$ are explicitly calculated using the formulas in (37) and (38). The calculations are somewhat tedious to do by hand, but since the formulas are explicit one can usually compute the parameters very quickly with technology.

[5] See [32] and [17] for the up to date account of the known results.
[6] $(X_n)_{n\geq 0}$ is said to be strongly transient under P_η if it is transient and $E_\eta[R | R < \infty] < \infty$, where $R := \inf\{n \geq 1 : X_n = X_0\}$.

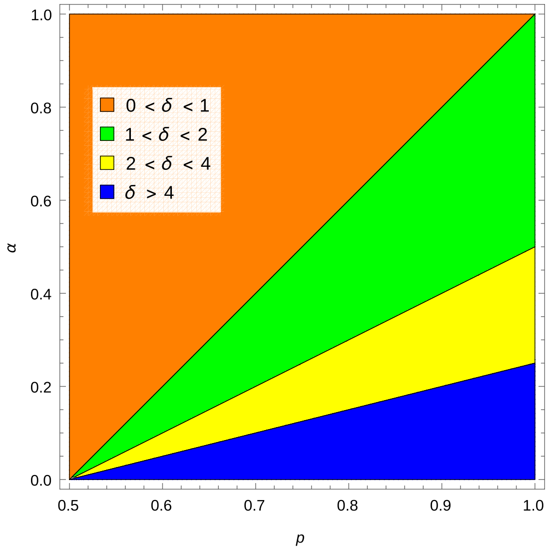


Fig. 1. A plot showing the regions for different types of behavior occurring for cookie environments as in Example 1.14.

Example 1.13 (Periodic cookie stacks). *In the setting of Example 1.1,*

$$\delta = \frac{\sum_{j=1}^N \sum_{i=1}^j (1-p(j))(2p(i)-1)}{2\sum_{j=1}^N p(j)(1-p(j))} \quad \textit{and} \quad \tilde{\delta} = \frac{\sum_{j=1}^N \sum_{i=1}^j p(j)(1-2p(i))}{2\sum_{j=1}^N p(j)(1-p(j))}.$$

For this model, the characterization of recurrence and transience was proved previously in [27] *but the results in Theorems 1.7 and 1.8 are new.*

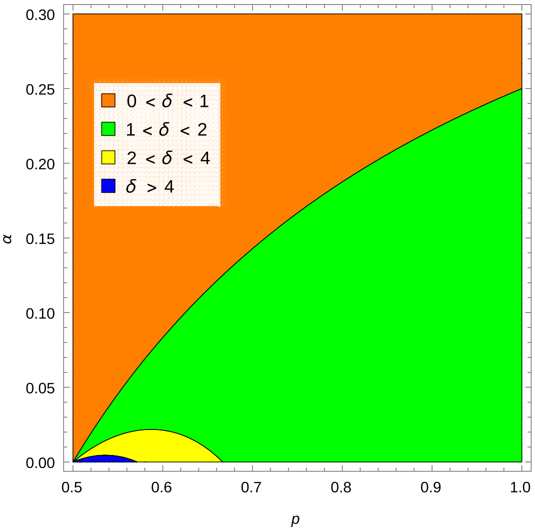
Example 1.14 (Cookie stacks of geometric height). *In the setting of Example 1.2 one obtains that* $\delta = -\tilde{\delta} = (2p(1)-1)/\alpha$. *Note that* $\mathbb{E}_1[\sum_{j=1}^\infty (2\omega_0(j)-1)] = (2p(1)-1)/\alpha$ *as well; this agrees with the criteria for recurrence/transience for this example that can be obtained from* [42]. *Previously, there were no known results regarding ballisticity or limiting distributions for this example. A plot showing the phase transitions for transience, ballisticity, and a CLT as a function of the parameters* α *and* $p(1)$ *for this example is given in Figure 1.*

Example 1.15 (Two-type, critical). *Let* $K = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$ *for some* $\alpha \in (0,1)$ *and let* $\mathbf{p} = (p, 1-p)^t$ *for some* $p > 1/2$. *In this case, if we use the initial condition* $\eta = (1,0)$ *then a calculation done with Mathematica yields*

$$\delta = \delta\left(\begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \begin{pmatrix} p \\ 1-p \end{pmatrix}, (1,0)\right) = \frac{(2p-1)((2\alpha-1)p-\alpha)}{4(2\alpha-1)(p-1)p+\alpha-1}. \tag{2}$$

Note that if $\alpha \in (0, 1/4)$ *then the parameter* δ *is non-monotone in* p. *In fact, for* $\alpha < \frac{14-3\sqrt{21}}{56} \approx 0.00450487$ *the parameter* δ *starts at* 0, *increases to a value larger than* 4, *and then decreases to* 1 *as* p *ranges from* 1/2 *to* 1 *(See Figure 2).*

Note also that the case $\alpha = 0$ corresponds to a classical simple random walk which steps to the right with probability $p > 1/2$ on each step. However, taking $\alpha = 0$ in (2) gives $\delta = \frac{p}{2p-1}$ and thus the results of Theorems 1.7 and 1.8 clearly do not hold when $\alpha = 0$ (this is not a contradiction since the transition matrix K is the identity matrix when $\alpha = 0$ and there is more than one irreducible closed set). The case $\alpha = 1$ gives periodic cookie stacks with period 2 and the formula (2) gives $\delta = \frac{(2p-1)(1-p)}{4p(1-p)}$ which agrees with the formula obtained in [27], see Example 1.13.



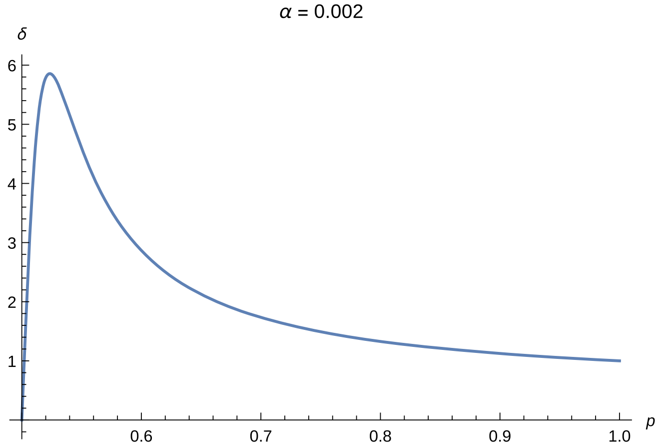


Fig. 2. The left plot shows the regions for different types of behavior occurring for cookie environments as in Example 1.15. The right plot shows the value of δ as a function of $p \in (1/2, 1)$ when $\alpha = 0.002$.

1.5. *Structure of the paper*

In Section 2 we introduce our main tool, forward and backward BLP, and show how the main results in the critical case (Theorems 1.6–1.8) can be deduced from Theorems 2.6 and 2.7 about the behavior of the tails of the lifetime and the total progeny over a lifetime of these BLP. Much of the remainder of the paper therefore is focused on the analysis of these BLPs, with the goal of proving Theorems 2.6 and 2.7. Section 3 discusses the asymptotics of the mean and variance of the forward BLP. Section 4 extends the results of the previous section to the backward BLP and establishes key relationships between parameters of the BLPs; in particular, in Section 4 we give explicit formulas for δ and $\tilde{\delta}$. In Section 5 we treat the non-critical case ($\bar{p} \neq 1/2$) deriving Theorem 1.5. Using the formulas for the parameters of the BLPs derived in the previous sections, we show how the BLPs in the non-critical case $\bar{p} > 1/2$ can be coupled with critical BLPs for which the parameter δ is arbitrarily large. From this coupling we then show that the conclusions of Theorem 1.5 can be derived from Theorems 2.6 and 2.7 in the same way as the proofs of Theorems 1.6–1.8 were obtained when $\delta > 4$. Finally, in Sections 6 and 7 we discuss proofs of Theorems 2.6 and 2.7. The crucial step here is showing that the BLPs have scaling limits that are squared Bessel processes. The generalized dimensions of these squared Bessel processes depend on the asymptotics of the mean and variance of the BLPs computed in Sections 3 and 4. This provides the connection of the parameters δ and $\tilde{\delta}$ defined in Section 4 with the tail asymptotics exponents in Theorems 2.6 and 2.7.

2. The associated branching-like processes

In this section, we introduce two BLPs that are naturally associated with the ERW and will prove the main theorems for the critical case assuming the necessary results about these BLP.

Given a cookie environment ω, we will expand the measure P_ω to include an independent family of Bernoulli random variables $\{\xi_x(j)\}_{x\in\mathbb{Z}, j\geq 1}$ such that $\xi_x(j) \sim \text{Bernoulli}(\omega_x(j))$. The ERW can then be constructed from the $\xi_x(j)$ as follows: if $X_n = x$ and $\sum_{k=0}^n \mathbf{1}_{\{X_k=x\}} = j$, then $X_{n+1} = X_n + 2\xi_x(j) - 1$.

Remark 2.1. *The arrow systems construction of the ERW introduced by Holmes and Salisbury* [18] *is very similar to the above coin toss construction*: *compare the last equation with* [17, *p.* 2, *above Theorem* 1.1]. *The distinction lies in the emphasis of the arrow approach on combinatorial results obtained by coupling of arrow systems rather than probability measures.*

We now show how the family of Bernoulli random variables $\{\xi_x(j)\}$ can also be used to construct the associated forward and backward BLP.

2.1. *The forward branching-like process*

The excursions of a random walk to the right of the origin induce a natural tree-like structure on the right-directed edge local times of the walk. That is, for $x \ge 1$ jumps from x to $x+1$ can be thought of "descendants" of previous jumps from $x-1$ to x. To be precise, if we set $\gamma_0 = 0$ and for $n \ge 1$ and $x \ge 0$ let

$$\gamma_n = \inf\{k > \gamma_{n-1} : X_{k-1} > 0 \text{ and } X_k = 0\} \quad \text{and} \quad \mathcal{E}_x^n = \sum_{k=0}^{\gamma_n - 1} \mathbf{1}_{\{X_k = x, X_{k+1} = x+1\}},$$

then γ_n is the time of the nth return to the origin from the right and $\mathcal{E}_x^n$ is the number of times the random walk has traversed the directed edge from x to $x+1$ by time γ_n (note that γ_n can be infinite if the walk is transient to the right or left).

If the random walk makes n excursions to the right ($\gamma_n < \infty$), then the directed edge local times $\mathcal{E}_x^n$ can be computed from the Bernoulli random variables. If $\mathcal{E}_{x-1}^n = m$, then since the steps of the random walk are ± 1 it follows that the walk makes m jumps to the left from x by time γ_n. Thus, $\mathcal{E}_x^n$ is the number of jumps that the random walk makes to the right from x before the mth jump to the left. If we refer to a Bernoulli random variable $\xi_x(j)$ as a "success" if $\xi_x(j) = 1$ and a "failure" if $\xi_x(j) = 0$, then $\mathcal{E}_x^n$ is the number of successes in the Bernoulli sequence $\xi_x = \{\xi_x(j)\}_{j \ge 1}$ before the mth failure. More precisely, introducing the notation

$$S_m^x = \inf\left\{k \ge 0 : \sum_{j=1}^{k+m} \bigl(1 - \xi_x(j)\bigr) = m\right\},$$

we have that if $\gamma_n < \infty$ and $\mathcal{E}_{x-1}^n = m$ then $\mathcal{E}_x^n = S_m^x$. Note that $S_0^x = 0$ by the convention that an empty sum is equal to zero. Also, let $G_m^x = S_m^x - S_{m-1}^x$ so that G_m^x is the number of successes between the $(m-1)$st failure and the mth failure in the Bernoulli sequence ξ_x.

With the above directed edge process $\mathcal{E}_x^n$ as motivation, we define the forward BLP started at $y \ge 1$ by

$$U_0 = y \quad \text{and} \quad U_i = S_{U_{i-1}}^i = \sum_{m=1}^{U_{i-1}} G_m^i \quad \text{for } i \ge 1. \tag{3}$$

We will use the notation $P_\omega^{U,y}(\cdot)$ and $P_\eta^{U,y}(\cdot) = \mathbb{E}_\eta[P_\omega^{U,y}(\cdot)]$ for the quenched and averaged distributions, respectively, of the forward BLP started at y. Note that under the quenched measure $P_\omega^{U,y}$ the forward BLP is a time-inhomogeneous Markov chain, but since the cookie environment is assumed to be spatially i.i.d. the forward BLP is a time-homogeneous Markov chain under the averaged measure $P_\eta^{U,y}$ for any initial distribution η.

The following Lemma summarizes the connection of the forward BLP with the directed edge local times of the random walk.

Lemma 2.2. *If $U_0 = n \ge 1$, then $U_i \ge \mathcal{E}_i^n$ for all $i \ge 0$. Moreover, on the event $\{\gamma_n < \infty\}$ we have $U_i = \mathcal{E}_i^n$ for all $i \ge 0$ if $U_0 = n$.*

The equality $U_i = \mathcal{E}_i^n$ on the event $\{\gamma_n < \infty\}$ was described above. For the proof of the general inequality $U_i \ge \mathcal{E}_i^n$ we refer the reader to [25, Section 4] or [32, Lemma 2.1].

Remark 2.3. *Note that in the case of classical simple random walks (i.e., $\omega_x(j) \equiv p \in (0,1)$) $\{G_m^i\}_{i \in \mathbb{Z}, m \ge 1}$ is a sequence of i.i.d.* Geometric$(1-p)$ *random variables. In this case, it is clear from* (3) *that U_i is a branching process with* Geometric$(1-p)$ *offspring distribution. For ERWs with boundedly many cookies per site the G_m^i are i.i.d.* Geometric$(1/2)$ *for all m sufficiently large and thus one can interpret U_i as a branching process with (random)* *migration* (*cf.* [25], *or more explicitly but in the context of the backward BLP see* [3]). *However, in the more general setup of the current paper one can no longer interpret* U_i *as a branching process and so we simply refer to* U_i *as a "branching-like" process.*

Remark 2.4. *One can also obtain a similar BLP which is related to the excursions of the ERW to the left of the origin. Clearly this BLP would have the same law as the forward BLP* U_i *defined here but with* $\mathbf{p}$ *replaced by* $\mathbf{1}-\mathbf{p}$.

2.2. *The backward branching-like process*

The backward BLP is related to the random walk through edge local times. However, the backward BLP is related to the local times of the *left* directed edges when the random walk first reaches a fixed point to the right of the origin. To be precise,

$$\mathcal{D}_n^x = \sum_{k=0}^{T_n-1} \mathbf{1}_{\{X_k=x, X_{k+1}=x-1\}}, \quad n \ge 1, x < n,$$

be the number of steps to the left from x before time T_n (recall that $T_n = \inf\{k \ge 0 : X_k = n\}$ is the hitting time of n by the ERW). On the event $\{T_n < \infty\}$, the sequence of directed edge local times $\{\mathcal{D}_n^x\}_{x \le n}$ also have a branching-like structure. Jumps to the left from $x+1$ before time T_n give rise to subsequent jumps to the left from x before time T_n. However, one important difference with the forward BLP should be noted in that not all jumps to the left from x are "descendants" of jumps to the left from $x+1$. In particular, for $x \ge 0$ the random walk can jump to the left from x before ever jumping from x to $x+1$.

As with $\mathcal{E}_x^n$ above, the directed edge process $\mathcal{D}_n^x$ can be computed from the Bernoulli random variables $\xi_x(j)$. In particular, if

$$F_m^x = \inf\left\{k \ge 0 : \sum_{j=1}^{k+m} \xi_x(j) = m\right\}$$

denotes the number of failures in the Bernoulli sequence $\xi_x = \{\xi_x(j)\}_{j\ge1}$ before the mth success then it is easy to see that if $T_n < \infty$ then $\mathcal{D}_n^n = 0$ and

$$\text{if } \mathcal{D}_n^{x+1} = m \quad \text{then } \mathcal{D}_n^x = \begin{cases} F_{m+1}^x, & x \ge 0, \\ F_m^x, & x < 0, \end{cases} \quad \text{for } x < n. \tag{4}$$

Indeed, if $x \ge 0$ and there are m jumps from $x+1$ to x before time T_n then there must be $m+1$ jumps from x to $x+1$ (the initial jump from x to $x+1$ plus m more jumps which can be paired with a prior jump from $x+1$ to x). Thus, from the construction of the random walk via the Bernoulli random variables $\xi_x(j)$ above, it follows that the number of jumps from x to $x-1$ before time T_n is the number of failures before the $(m+1)$th success in the Bernoulli sequence ξ_x. The explanation of (4) when $x < 0$ is similar, with the exception that all jumps to the right from x can be paired with a prior jump to the left from $x+1$.

Again with the directed edge local time process as motivation, we define the backward BLP started at $y \ge 0$ by

$$V_0 = y \quad \text{and} \quad V_i = F_{V_{i-1}+1}^i \quad \text{for } i \ge 1. \tag{5}$$

We will use $P_\omega^{V,y}$ and $P_\eta^{V,y}$ to denote the quenched and averaged laws of the backward branching process started at $y \ge 0$. As with the forward BLP, V_i is a time-inhomogeneous Markov chain under the quenched measure and a time-homogeneous Markov chain under the averaged measure.

Lemma 2.5. *If* $n \ge 1$ *and* $P_\eta(T_n < \infty) = 1$, *then the sequence* $(\mathcal{D}_n^n, \mathcal{D}_n^{n-1}, \dots, \mathcal{D}_n^1, \mathcal{D}_n^0)$ *has the same distribution under* P_η *as the sequence* $(V_0, V_1, \dots, V_{n-1}, V_n)$ *under the measure* $P_\eta^{V,0}$.

Proof. For $n \geq 1$ let $\{\xi_x^{(n)}(j)\}_{x\in\mathbb{Z}, j\geq 1}$ be the family of Bernoulli random variables given by $\xi_x^{(n)}(j) = \xi_{n-x}(j)$ for $x \in \mathbb{Z}$ and $j \geq 1$, and let $V_i^{(n)}$ be the backward BLP started at $V_0^{(n)} = 0$ but defined using the Bernoulli family $\{\xi_x^{(n)}(j)\}$ in place of $\{\xi_x(j)\}$. Then, it is clear from (4) and (5) on the event $\{T_n < \infty\}$ that $\mathcal{D}_n^{n-i} = V_i^{(n)}$ for $i = 0, 1, \dots, n$. Finally, since the cookie environments are spatially i.i.d., it follows that $\{\xi_x^{(n)}(j)\}_{x\in\mathbb{Z}, j\geq 1}$ and $\{\xi_x(j)\}_{x\in\mathbb{Z}, j\geq 1}$ have the same distribution under the averaged measure $P_\eta = \mathbb{E}_\eta[P_\omega(\cdot)]$. □

2.3. *Proofs of the main results in the critical case*

Here and throughout the remainder of the paper, we will use the following notation for hitting times of stochastic processes. If $\{Z_j\}_{j\geq 0}$ is a stochastic process, then for any $x \in \mathbb{R}$ let σ_x^Z and τ_x^Z be the hitting times

$$\sigma_x^Z := \inf\{j > 0 : Z_j \leq x\} \quad \text{and} \quad \tau_x^Z := \inf\{j \geq 0 : Z_j \geq x\}.$$

Usually the stochastic process Z will be either the forward or backward BLP, though occasionally we will also use this notation for other processes.

The analysis of the forward and backward BLP is key to the proofs of all the main results in this paper. In particular, all of the results in the critical case $\bar{p} = 1/2$ (Theorems 1.6–1.8) will follow from the following two theorems.

Theorem 2.6. *Let $\bar{p} = 1/2$ and let δ be given by (38). If U is the forward BLP defined in (3), then*:

- *If $\delta > 1$ then $P_\eta^{U,y}(\sigma_0^U = \infty) > 0$ for all $y \geq 1$.*
- *If $\delta \leq 1$ then for any $y \geq 1$ there are positive constants $c_1 = c_1(y, \eta)$ and $c_2 = c_2(y, \eta)$ such that*

$$\lim_{n\to\infty} n^{1-\delta} P_\eta^{U,y}\big(\sigma_0^U > n\big) = c_1 \quad \text{and} \quad \lim_{n\to\infty} n^{(1-\delta)/2} P_\eta^{U,y}\left(\sum_{j=0}^{\sigma_0^U - 1} U_j > n\right) = c_2, \tag{6}$$

where for $\delta = 1$ we replace n^0 with $\ln n$.

Theorem 2.7. *Let $\bar{p} = 1/2$ and let δ be given by (38). If V is the backward BLP defined in (5), then*:

- *If $\delta < 0$ then $P_\eta^{V,y}(\sigma_0^V = \infty) > 0$ for all $y \geq 0$.*
- *If $\delta \geq 0$ then for any $y \geq 0$ there are positive constants $c_3 = c_3(y, \eta)$ and $c_4 = c_4(y, \eta)$ such that*

$$\lim_{n\to\infty} n^{\delta} P_\eta^{V,y}\big(\sigma_0^V > n\big) = c_3 \quad \text{and} \quad \lim_{n\to\infty} n^{\delta/2} P_\eta^{V,y}\left(\sum_{j=0}^{\sigma_0^V - 1} V_j > n\right) = c_4, \tag{7}$$

where for $\delta = 0$ we replace n^0 with $\ln n$.

Remark 2.8. *Analogous theorems are known for the ERW with bounded number of cookies per site*, [23, *Theorems* 2.1, 2.2], [9, *Theorem* 1.3].

We will give the proofs of Theorems 2.6 and 2.7 in Sections 6 and 7 below, but first we will show how they are used to prove Theorems 1.6–1.8.

Proof of Theorem 1.6. For any cookie environment ω, let ω^+ be the modified cookie environment in which $\omega_x^+(j) = \omega_x(j)$ for $x \neq 0$, $j \geq 1$ and with $\omega_0^+(j) = 1$ for all $j \geq 1$ (that is, in the cookie environment ω^+ the walk steps to the right after every visit to the origin). Recall that γ_n is the time of the nth return to the origin. Lemma 2.2 implies that

$$P_{\omega^+}(\gamma_n < \infty) = P_{\omega^+}\big(\mathcal{E}_n^x = 0, \text{ for all } x \text{ sufficiently large}\big) \leq P_\omega^{U,n}\big(\sigma_0^U < \infty\big). \tag{8}$$

(Note that in the last probability on the right we can change the cookie environment from ω^+ to ω since the forward branching process is generated using the Bernoulli random variables $\xi_x(j)$ with $x \geq 1$.) If instead $\gamma_n = \infty$, then since the $\omega_x(j)$ are uniformly bounded away from 0 and 1 the walk cannot stay bounded for the first n excursions to the right. Therefore,

$$P_{\omega^+}(\gamma_n = \infty) = P_{\omega^+}\bigl(\mathcal{E}_n^x \ge 1, \forall x \ge 1\bigr) \le P_\omega^{U,n}\bigl(\sigma_0^U = \infty\bigr), \tag{9}$$

where the last inequality follows from Lemma 2.2. Combining (8) and (9) we can conclude that

$$P_{\omega^+}(\gamma_n < \infty) = P_\omega^{U,n}\bigl(\sigma_0^U < \infty\bigr), \quad \forall n \ge 1. \tag{10}$$

Suppose that $\delta \le 1$. Then it follows from (10) and Theorem 2.6 that $E_\eta[P_{\omega^+}(\gamma_n < \infty)] = 1$ for all $n \ge 1$. That is, with probability one every excursion of the ERW to the right of the origin will eventually return to the origin. Similarly, if $\tilde{\delta} \le 1$ then all excursions to the left of the origin eventually return to the origin. Therefore, if $\delta, \tilde{\delta} \le 1$ then all excursions from the origin are finite and so the walk returns to the origin infinitely many times. Since all $\omega_x(j)$ are uniformly bounded away from 0 and 1 this implies that the walk visits every site infinitely often.

If instead $\delta > 1$, then (10) and Theorem 2.6 imply that $E_\eta[P_{\omega^+}(\gamma_1 = \infty)] > 0$, and thus

$$0 < E_\eta\bigl[\omega_0(1)\bigr]E_\eta\bigl[P_{\omega^+}(\gamma_1 = \infty)\bigr] \le P_\eta(X_n \ge 1, n \ge 1) \le P_\eta\Bigl(\lim_{n\to\infty} X_n = \infty\Bigr),$$

where the last inequality again follows from the fact that the $\omega_x(j)$ are uniformly bounded away from 0 and 1. Finally, we can conclude by Theorem 1.4 that $\delta > 1$ implies that the walk is transient to the right with probability 1. A similar argument shows that $\tilde{\delta} > 1$ implies that the walk is transient to the left. □

Proof of Theorem 1.7. Since the limiting speed v_0 exists by Theorem 1.4(ii), if the walk is recurrent then the speed must be $v_0 = 0$. Thus, we need only to consider the case where the walk is transient. First, assume that the walk is transient to the right ($\delta > 1$). Since $T_n/n = T_n/X_{T_n}$, the existence of the limiting speed v_0 implies that $\lim_{n\to\infty} T_n/n = 1/v_0$, where we use the convention $1/0 = \infty$. It is easy to see that $T_n = n + 2\sum_{x<n} \mathcal{D}_n^x$ for every $n \ge 1$. Since the walk is transient to the right,

$$\lim_{n\to\infty} 2\sum_{x<0} \mathcal{D}_n^x \le \sum_{k=0}^{\infty} \mathbf{1}_{\{X_k \le -1\}} < \infty, \tag{11}$$

and thus

$$\frac{1}{v_0} = \lim_{n\to\infty} \frac{T_n}{n} = \lim_{n\to\infty} \frac{1}{n}\Bigl(n + 2\sum_{x<n} \mathcal{D}_n^x\Bigr) = 1 + \lim_{n\to\infty} \frac{2}{n}\sum_{x=0}^{n-1} \mathcal{D}_n^x.$$

It follows from Lemma 2.5 that $n^{-1}\sum_{x=0}^{n-1} \mathcal{D}_n^x$ has the same distribution as $n^{-1}\sum_{i=1}^{n} V_i$ started with $V_0 = 0$, and standard Markov chain arguments imply that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} V_i = \frac{E_\eta^{V,0}[\sum_{i=0}^{\sigma_0^V-1} V_i]}{E_\eta^{V,0}[\sigma_0^V]}, \qquad P_\eta^{V,0}\text{-a.s.}$$

Thus, we can conclude that

$$v_0 = \left(1 + 2\frac{E_\eta^{V,0}[\sum_{i=0}^{\sigma_0^V-1} V_i]}{E_\eta^{V,0}[\sigma_0^V]}\right)^{-1}.$$

Theorem 2.7 implies that the $E_\eta^{V,0}[\sigma_0^V] < \infty$ (recall that $\delta > 1$ since the walk is transient to the right) and that $E_\eta^{V,0}[\sum_{i=0}^{\sigma_0^V-1} V_i] < \infty$ if and only if $\delta > 2$. Thus we can conclude that $v_0 > 0 \iff \delta > 2$. Again, a symmetric argument for random walks that are transient to the left shows that $v_0 < 0 \iff \tilde{\delta} > 2$. □

Proof of Theorem 1.8. The proofs of the limiting distributions in Theorem 1.8 relies on the connection of the hitting times T_n with the backward BLP V_i from Lemma 2.5 and the tail asymptotics for the backward BLP in (7). We will give a brief sketch of the general argument here and will give full details in the case $\delta = 2$ in Appendix B. We will refer the reader to previous papers for the details in all other cases.

For the limiting distributions of the hitting times T_n, recall that

$$T_n = n + 2\sum_{x=0}^{n-1} \mathcal{D}_n^x + 2\sum_{x<0} \mathcal{D}_n^x. \tag{12}$$

It follows from (11) that the third term on the right has a finite limit as $n \to \infty$, P_η-a.s., and therefore it is enough to prove a limiting distribution for the first two terms on the right side of (12). By Lemma 2.5 this is equivalent to proving a limiting distribution for $n + 2\sum_{i=1}^n V_i$ under the measure $P_\eta^{V,0}$. The proof of the limiting distribution for the partial sums of the BLP relies on the regeneration structure of the process V_i. Let r_k denote the time of the kth return of the backward BLP to zero. That is,

$$r_0 = 0 \quad \text{and} \quad r_k = \inf\{i > r_{k-1} : V_i = 0\}.$$

Also, let $W_k = \sum_{i=r_{k-1}}^{r_k - 1} V_i$. Note that $\{(r_k - r_{k-1}, W_k)\}_{k\ge 1}$ is an i.i.d. sequence under the measure $P_\eta^{V,0}$ and that Theorem 2.7 implies that r_1 and W_1 are in the domains of attraction of totally asymmetric stable distributions of index $\min\{\delta, 2\}$ and $\min\{\delta/2, 2\}$, respectively. From this, the limiting distributions for $n + 2\sum_{i=1}^n V_i$ are standard. For the details of the arguments in cases (i)–(v) we give the following references.

- $\delta \in (1, 2)$: See [4], pages 847–849.
- $\delta = 2$: See Appendix B.
- $\delta \in (2, 4]$: See Section 9 in [23].
- $\delta > 4$: By (7) the second moment of the random variable $\sum_{j=0}^{\sigma_0^V - 1} V_j$ is finite. Thus, this is a classical case covered by the standard CLT for the Markov chain V; see, for example, [6, I.16, Theorem 1].[7]

Finally, to obtain the limiting distributions for the position X_n of the ERW from the limiting distributions of the hitting times T_n, we use the fact that

$$\{T_m > n\} \subset \{X_n < m\} \subset \{T_{m+r} > n\} \cup \Big\{\inf_{k\ge T_{m+r}} X_k < m\Big\}, \tag{13}$$

for any $m, n, r \ge 1$. The key then is to control the probability of the last event on the right. To this end, it was shown in [31, Lemma 6.1] that the tail asymptotics for r_1 in (7) imply that

$$P_\eta\Big(\inf_{k\ge T_{m+r}} X_k < m\Big) \le C r^{1-\delta}, \quad \forall m, r \ge 1. \tag{14}$$

Again, for the details of how to use (13) and (14) to obtain the limiting distributions for X_n, see the references given above. □

3. Mean and variance of the forward BLP

Many of the calculations below are simplified using matrix notation.

- $\mathbf{1}$ will denote a column vector of all ones, and $\mathbf{p}$ denotes a column vector with ith entry $p(i)$.
- I is the identity matrix.
- D_p will denote a diagonal matrix with ith diagonal entry $p(i)$. Similarly $D_{1-p} = I - D_p$ is the diagonal matrix with ith diagonal entry $1 - p(i)$.

[7] For the second statement of (v) see also [25], proof of Theorem 3 and Section 6. The argument is based on the result due to A.-S. Sznitman [37, Theorem 4.1] and gives the functional CLT for the position of the walk.

- Recall that $\bar{p} = \mu \cdot \mathbf{p}$, where μ is the stationary distribution for the environment Markov chain.

In this section and the next section, we will be concerned with a single increment of the BLP. Thus, we will only need to consider the environment at any fixed site $x \in \mathbb{Z}$. Therefore, in this section we will fix x and suppress the sub/super-script x for a less cumbersome notation. For instance, we will write $R_j = R_j^x$ for the Markov chain which generates the environment at x and the Bernoulli sequence is denoted $\{\xi(j)\}_{j\ge 1} = \{\xi_x(j)\}_{j\ge 1}$.

3.1. *Mean*

Proposition 3.1. *For every distribution η on $\mathcal{R}$*

$$\lim_{n\to\infty} \frac{E_\eta^{U,n}[U_1]}{n} = \frac{\bar{p}}{1-\bar{p}} =: \lambda.$$

Moreover, there exist constants $c_5, c_6 > 0$ such that for any $n \in \mathbb{N}$ and any distribution η on $\mathcal{R}$

$$\left| E_\eta^{U,n}[U_1] - \lambda n - \eta \cdot \mathbf{r} \right| \le c_5 e^{-c_6 n}, \tag{15}$$

where

$$\mathbf{r} = \mathbf{r}(\mathbf{p}, K) = \left(I - K + \frac{(\mathbf{1}-\mathbf{p})\mu D_{1-p} K}{1-\bar{p}} \right)^{-1} \mathbf{p} - \lambda \mathbf{1}. \tag{16}$$

Proof. Recall that S_m is the number of successes before the mth failure in the Bernoulli sequence $\xi = \{\xi(j)\}_{j\ge 1}$, and let $G_m = S_m - S_{m-1}$ be the number of successes between the $(m-1)$st and the mth failure in the Bernoulli sequence ξ. With this notation it follows from the construction of the forward BLP in Section 2 that

$$E_\eta^{U,n}[U_1] = \sum_{m=1}^{n} E_\eta[G_m]. \tag{17}$$

To compute $E_\eta[G_m]$ it helps to keep track of some additional information. Let $I_0 = R_1$ and for any $m \ge 1$ let I_m be defined by

$$I_m = R_{S_m+m+1}.$$

Note that $S_m + m$ is the number of Bernoulli trials needed to obtain m failures. Therefore, $I_m = i$ if the next Bernoulli random variable after the mth failure has success probability $p(i)$. Since

$$P_\eta\big(\{\xi(j)\}_{j\ge S_m+m+1} \in \cdot \,|\, \sigma(\xi(j), j \le S_m + m + 1)\big) = P_{I_m}\big(\{\xi(j)\}_{j\ge 1} \in \cdot\big),$$

it follows that $\{I_m\}_{m\ge 0}$ is a Markov chain on the state space $\mathcal{R}$. By the ellipticity assumption, $p : \mathcal{R} \to (0,1)$, the Markov chain $\{I_m\}_{m\ge 0}$ has the same unique closed irreducible subset $\mathcal{R}_0$ as the Markov chain $\{R_j\}_{j\ge 1}$. Therefore, $\{I_m\}_{m\ge 0}$ has a unique stationary distribution π. While we did not assume any aperiodicity for $\{R_j\}_{j\ge 0}$, the ellipticity assumption implies that $\{I_m\}_{m\ge 0}$ is aperiodic, and since $\mathcal{R}$ is finite the convergence to stationarity is exponentially fast: if Π is the matrix of transition probabilities for the Markov chain I_m, then there exist constants $c_7, c_8 > 0$ such that

$$\sup_\eta \left\| \eta \Pi^n - \pi \right\|_\infty \le c_7 e^{-c_8 n}, \quad \forall n \ge 1, \tag{18}$$

where the supremum on the left is over all probability measures η on $\mathcal{R}$.

Let $\mathbf{g}$ denote the column vector of length N with ith entry $g(i) = E_i[G_1]$. Note that by the comparison with a geometric random variable with parameter $p_{\max} := \max_{i\in\mathcal{R}} p(i) < 1$ we see that $g(i) < p_{\max}/(1-p_{\max}) < \infty$ for every $i \in \mathcal{R}$. Then, since $E_\eta[G_m] = E_\eta[g(I_{m-1})]$ it follows from an ergodic theorem for finite state Markov chains that

$$\lim_{n\to\infty} \frac{1}{n} E_\eta^{U,n}[U_1] = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} E_\eta[G_m] = E_\pi\bigl[g(I_0)\bigr] = \sum_{i\in\mathcal{R}} \pi(i)g(i) = \pi \cdot \mathbf{g}.$$

Therefore, to prove the first part of the lemma we need to show that $\pi \cdot \mathbf{g} = \lambda$. The following lemma accomplishes this task. It also provides useful information about $\{I_m\}_{m\geq 0}$ which we need in the rest of this section.

Lemma 3.2. *The sequence $\{I_m\}_{m\geq 0}$ is a Markov chain with transition probabilities given by the matrix*

$$\Pi = (I - D_p K)^{-1} D_{1-p} K,$$

and with a unique stationary distribution π given by

$$\pi = \frac{\mu(I - D_p K)}{1 - \bar{p}} = (1 + \lambda)\mu(I - D_p K). \tag{19}$$

Moreover,

$$\mathbf{g} = (I - D_p K)^{-1}\mathbf{p} \quad \textit{and} \quad \pi \cdot \mathbf{g} = \lambda. \tag{20}$$

Remark 3.3. *In the first formula for π in (19), the multiplicative factor $1/(1 - \bar{p})$ is needed for the entries of π to sum to 1 since $\mu(I - D_p K)\mathbf{1} = \mu(\mathbf{1} - \mathbf{p}) = 1 - \bar{p}$. The second formula for π in (19) is equivalent since $\lambda = \frac{\bar{p}}{1-\bar{p}}$ implies that $\frac{1}{1-\bar{p}} = 1 + \lambda$.*

Proof. To compute the transition probabilities, for $k \geq 0$ let M_k be the $N \times N$ matrix with entries

$$M_k\bigl(i, i'\bigr) = P_i\bigl(G_1 = k, I_1 = i'\bigr).$$

Obviously, $M_0(i, i') = (1 - p_i)K_{i,i'} = (D_{1-p}K)_{i,i'}$ and for $k \geq 1$ by conditioning on the value of R_2 we obtain that

$$M_k\bigl(i, i'\bigr) = \sum_{\ell=1}^{N} p_i K_{i,\ell} M_{k-1}\bigl(\ell, i'\bigr).$$

That is, $M_0 = D_{1-p}K$ and $M_k = D_p K M_{k-1}$ for $k \geq 1$. Combining these we get that $M_k = (D_p K)^k D_{1-p}K$. Since $\Pi(i, i') = \sum_{k\geq 0} M_k(i, i')$, we have that

$$\Pi = \sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty} (D_p K)^k D_{1-p} K = (I - D_p K)^{-1} D_{1-p} K.$$

(Note that since $D_p K$ is a matrix with non-negative entries and with ith row sum equal to $p(i) < 1$, the Perron-Frobenius theorem implies that all eigenvalues of $D_p K$ have absolute value strictly less than one, and thus $I - D_p K$ is invertible.) It is easy to check that $\pi\Pi = \pi$ by noting that $\mu K = \mu$ and, thus,

$$\mu D_{1-p} K = \mu(I - D_p)K = \mu(I - D_p K). \tag{21}$$

Finally, we give a formula for $\mathbf{g}$. For $k \geq 1$ one easily sees that

$$P_i(G_1 \geq k) = \sum_{i_1\in\mathcal{R}} p(i) K_{i,i_1} P_{i_1}(G_1 \geq k - 1).$$

Iterating this, we obtain

$$P_i(G_1 \geq k) = \sum_{i_1, i_2, \ldots, i_k \in \mathcal{R}} \big(p(i) K_{i,i_1}\big)\big(p(i_1) K_{i_1,i_2}\big) \cdots \big(p(i_{k-1}) K_{i_{k-1},i_k}\big) = e_i (D_p K)^k \mathbf{1}, \tag{22}$$

where in the last equality we use the notation e_i for the row vector with a one in the ith coordinate and zeros elsewhere. Therefore,

$$g(i) = E_i[G_1] = \sum_{k \geq 1} P_i(G_1 \geq k) = \sum_{k \geq 1} e_i (D_p K)^k \mathbf{1} = e_i (I - D_p K)^{-1} D_p K \mathbf{1} = e_i (I - D_p K)^{-1} \mathbf{p}$$

and we get (20) as claimed. From this and the formula for π in (19) it follows immediately that $\pi \cdot \mathbf{g} = \lambda$. □

In the critical case $\bar{p} = 1/2$, Lemma 3.2 and (21) give the following simpler formula for the stationary distribution π that will be useful below.

Corollary 3.4. *If $\bar{p} = 1/2$ then $\pi = 2\mu(I - D_p K) = 2\mu D_{1-p} K$.*

Thus far we have proved the first part of Proposition 3.1. Next we show the existence of a vector $\mathbf{r}$ such that (15) holds. To this end, for any $n \geq 1$ let $\mathbf{r}_n = (r_n(1), r_n(2), \ldots, r_n(N))^t$ be the column vector with ith entry

$$r_n(i) = E_i^{U,n}[U_1] - \lambda n = \sum_{m=1}^{n} \big(E_i[G_m] - \lambda\big).$$

Then

$$\mathbf{r}_n = \sum_{m=1}^{n} \big(\Pi^{m-1} \mathbf{g} - \mathbf{1}\pi \cdot \mathbf{g}\big) = \sum_{m=0}^{n-1} \big(\Pi^m - \mathbf{1}\pi\big) \mathbf{g},$$

where in the last equality the matrix $\mathbf{1}\pi$ is the $N \times N$ matrix with all rows equal to the vector π which is the stationary distribution for $\{I_m\}_{m \geq 0}$. It follows from (18) that the entries of $\Pi^m - \mathbf{1}\pi$ decrease exponentially in m, so that the sum in the last line converges as $n \to \infty$. That is,

$$\lim_{n \to \infty} \mathbf{r}_n = \mathbf{r} := \sum_{m=0}^{\infty} \big(\Pi^m - \mathbf{1}\pi\big) \mathbf{g}. \tag{23}$$

Moreover, since $E_\eta^{U,n}[U_1] - \lambda n = \eta \cdot \mathbf{r}_n$, then

$$\big|E_\eta^{U,n}[U_1] - \lambda n - \eta \cdot \mathbf{r}\big| = \big|\eta \cdot (\mathbf{r}_n - \mathbf{r})\big| \leq \sum_{m=n}^{\infty} \big\|\eta \Pi^m - \pi\big\|_\infty \|\mathbf{g}\|_1 \leq \frac{\|\mathbf{g}\|_1 c_7}{1 - e^{-c_8}} e^{-c_8 n},$$

where the last inequality follows from (18).

Finally, we give an explicit formula for the vector $\mathbf{r}$. To this end, since $\pi \Pi = \pi$ it follows that $(\mathbf{1}\pi)\Pi = \mathbf{1}\pi = \Pi(\mathbf{1}\pi)$ and thus $(\Pi^m - \mathbf{1}\pi) = (\Pi - \mathbf{1}\pi)^m$ for all $m \geq 1$. Therefore,

$$\mathbf{r} = (I - \mathbf{1}\pi)\mathbf{g} + \sum_{m=1}^{\infty} (\Pi - \mathbf{1}\pi)^m \mathbf{g} = \sum_{m=0}^{\infty} (\Pi - \mathbf{1}\pi)^m \mathbf{g} - (\pi \cdot \mathbf{g})\mathbf{1} = (I - \Pi + \mathbf{1}\pi)^{-1} \mathbf{g} - \lambda \mathbf{1}. \tag{24}$$

Substituting the expressions for π, Π, and $\mathbf{g}$ and simplifying we obtain (16). □

In closing this subsection, we note the following Corollary which will be of use later.

Corollary 3.5. $\pi \cdot \mathbf{r} = 0$.

Proof. Since π is the stationary distribution for the Markov chain with transition matrix Π it follows that $\pi(I - \Pi + \mathbf{1}\pi) = \pi$, or equivalently $\pi(I - \Pi + \mathbf{1}\pi)^{-1} = \pi$. Using the formula for $\mathbf{r}$ in (24), it follows that $\pi \cdot \mathbf{r} = \pi(I - \Pi + \mathbf{1}\pi)^{-1}\mathbf{g} - \lambda\pi \cdot \mathbf{1} = \pi \cdot \mathbf{g} - \lambda = 0$, where the last equality follows from (20). □

3.2. *Variance*

The main result of this subsection is the following proposition.

Proposition 3.6. *There exists a constant $c_9 > 0$ such that for any distribution η on $\mathcal{R}$*

$$\left|\frac{\mathrm{Var}_\eta(U_1|U_0 = n)}{n} - \nu\right| \le \frac{c_9}{n},$$

where the constant ν is given by the formula

$$\nu = \nu(\mathbf{p}, K) = \mathrm{Var}_\pi(G_1) + 2\sum_{k=1}^{\infty} \mathrm{Cov}_\pi(G_1, G_{1+k}) = (1+\lambda)(\lambda + 2\mu D_p K\mathbf{r}), \tag{25}$$

and $\mathbf{r}$ is the vector from Proposition 3.1. In particular, if $\bar{p} = 1/2$ then

$$\nu = 2 + 4\mu D_p K\mathbf{r} = 4\mu D_p K\bigl(I - K + 2(\mathbf{1} - \mathbf{p})\mu D_{1-p} K\bigr)^{-1}\mathbf{p}. \tag{26}$$

Proof. First note that for any measure η on $\mathcal{R}$,

$$\mathrm{Var}_\eta(U_1|U_0 = n) = \mathrm{Var}_\eta\left(\sum_{k=1}^{n} G_k\right) = \sum_{k=1}^{n} \mathrm{Var}_\eta(G_k) + 2\sum_{1\le k<\ell\le n} \mathrm{Cov}_\eta(G_k, G_\ell).$$

Let $\nu_n = \mathrm{Var}_\pi(U_1|U_0 = n) = \mathrm{Var}_\pi(\sum_{k=1}^n G_k)$. The proof will consist of three steps.

Step 1. Show that $|\nu_n - \mathrm{Var}_\eta(U_1|U_0 = n)|$ is bounded by a constant uniformly in n and η.
Step 2. Prove that there is a constant $C > 0$ such that $|\nu_n/n - \nu| \le C/n$.
Step 3. Calculate ν explicitly and show that (25) and (26) hold.

Step 1. For any $k \ge 0$ and $i \in \mathcal{R}$ let $v_k(i) = E_i[G_1 G_{1+k}]$ and

$$\mathbf{v}_k = \bigl(v_k(1), v_k(2), \ldots, v_k(N)\bigr)^t.$$

With this notation, we have that

$$\mathrm{Var}_\eta(G_k) = \eta\Pi^{k-1}\mathbf{v}_0 - \bigl(\eta\Pi^{k-1}\mathbf{g}\bigr)^2, \quad \forall k \ge 1, \tag{27}$$

and

$$\mathrm{Cov}_\eta(G_k, G_\ell) = \eta\Pi^{k-1}\mathbf{v}_{\ell-k} - \bigl(\eta\Pi^{k-1}\mathbf{g}\bigr)\bigl(\eta\Pi^{\ell-1}\mathbf{g}\bigr), \quad \forall 1 \le k < \ell. \tag{28}$$

Note that in the special case where η has distribution π these become

$$\mathrm{Var}_\pi(G_k) = \pi(\mathbf{v}_0 - \lambda\mathbf{g}), \quad \text{and} \quad \mathrm{Cov}_\pi(G_k, G_\ell) = \pi(\mathbf{v}_{\ell-k} - \lambda\mathbf{g}), \quad \text{for } 1 \le k < \ell, \tag{29}$$

since $\pi\Pi = \pi$ and $\pi \cdot \mathbf{g} = \lambda$. The following lemma is elementary.

Lemma 3.7. $\sup_{k\ge 0} \|\mathbf{v}_k\|_\infty < \infty$.

Proof. First of all, note that the Cauchy–Schwartz inequality implies that

$$E_i[G_1 G_{1+k}] \le \sqrt{E_i\big[G_1^2\big]E_i\big[G_{1+k}^2\big]} \le \max_i E_i\big[G_1^2\big] = \|\mathbf{v}_0\|_\infty,$$

and thus it is enough to prove that $E_i[G_1^2] < \infty$ for every $i \in \mathcal{R}$. The last inequality is obvious by comparison with a geometric random variable with parameter $p_{\max} = \max_{i\in\mathcal{R}} p(i) < 1$. □

Lemma 3.8. *There exist constants*, $c_{10}, c_{11} > 0$ *so that for any distribution* η *on* $\mathcal{R}$ *and any* $1 \le k < \ell$

$$\big|\mathrm{Var}_\eta(G_k) - \mathrm{Var}_\pi(G_k)\big| \le c_{10} e^{-c_{11}k} \quad \textit{and} \quad \big|\mathrm{Cov}_\eta(G_k, G_\ell) - \mathrm{Cov}_\pi(G_k, G_\ell)\big| \le c_{10} e^{-c_{11}\ell}. \tag{30}$$

Proof. The key observation is the exponential convergence to the stationary distribution π for the Markov chain $\{I_m\}_{m\ge 0}$ as noted in (18) above. From this, it follows easily that

$$\big|\eta\Pi^{k-1}\mathbf{v}_0 - \pi\cdot\mathbf{v}_0\big| \le c_7\|\mathbf{v}_0\|_1 e^{-c_8(k-1)} \quad \text{and} \quad \big|\eta\Pi^{k-1}\mathbf{g} - \pi\cdot\mathbf{g}\big| \le c_7\|\mathbf{g}\|_1 e^{-c_8(k-1)}.$$

The first inequality in (30) then follows easily from the above bounds and the representations for the variances in (27) and (29) taking into account the fact that $\eta\Pi^{k-1}$ is always a probability distribution so that $|\eta\Pi^{k-1}\mathbf{g}+\pi\cdot\mathbf{g}| \le 2\|\mathbf{g}\|_\infty$.

To obtain the bound on the difference of the covariance terms in (30), note that the representations in (28) and (29) imply that

$$\begin{aligned}\big|\mathrm{Cov}_\eta(G_k, G_\ell) - \mathrm{Cov}_\pi(G_k, G_\ell)\big| &\le \big|\eta\Pi^{k-1}\mathbf{v}_{\ell-k} - \big(\eta\Pi^{k-1}\mathbf{g}\big)\big(\eta\Pi^{\ell-1}\mathbf{g}\big) - \pi(\mathbf{v}_{\ell-k} - \lambda\mathbf{g})\big| \\ &\le \big|\big(\eta\Pi^{k-1} - \pi\big)(\mathbf{v}_{\ell-k} - \lambda\mathbf{g})\big| + \big|\eta\Pi^{k-1}\mathbf{g}\big|\big|\lambda - \eta\Pi^{\ell-1}\mathbf{g}\big| \\ &\le c_7 e^{-c_8(k-1)}\|\mathbf{v}_{\ell-k} - \lambda\mathbf{g}\|_1 + \|\mathbf{g}\|_\infty\|\mathbf{g}\|_1 c_7 e^{-c_8(\ell-1)},\end{aligned}$$

where the last inequality again follows from (18) and the fact that $\pi\cdot\mathbf{g} = \lambda$. Therefore, it will be enough to show that there exist constants $C, C' > 0$ such that

$$\|\mathbf{v}_k - \lambda\mathbf{g}\|_1 \le C\|\mathbf{g}\|_1^2 e^{-C'k}, \quad \forall k \ge 1. \tag{31}$$

To this end, note that by conditioning on (G_1, I_1) we get that

$$E_i[G_1 G_{k+1}] - \lambda E_i[G_1] = \sum_{i'\in\mathcal{R}}\sum_{n=0}^{\infty} n P_i\big(G_1 = n, I_1 = i'\big)\big(E_{i'}[G_k] - \lambda\big).$$

By (18) $\max_{i\in\mathcal{R}} |E_i[G_k] - \lambda| = \max_{i\in\mathcal{R}} |e_i\Pi^{k-1}\mathbf{g} - \pi\cdot\mathbf{g}| \le c_7\|\mathbf{g}\|_1 e^{-c_8(k-1)}$, and thus it follows that

$$\big|E_i[G_1 G_{k+1}] - \lambda E_i[G_1]\big| \le E_i[G_1]\big(c_7\|\mathbf{g}\|_1 e^{-c_8(k-1)}\big).$$

Since this gives a bound on each of the entries of $\mathbf{v}_k - \lambda\mathbf{g}$, the inequality in (31) follows. □

To complete step 1 in the proof of Proposition 3.6 we notice that Lemma 3.8 implies

$$\big|\nu_n - \mathrm{Var}_\eta(U_1|U_0 = n)\big| = \left|\mathrm{Var}_\pi\left(\sum_{k=1}^n G_k\right) - \mathrm{Var}_\eta\left(\sum_{k=1}^n G_k\right)\right| \le \sum_{k=1}^n c_{10}e^{-c_{11}k} + 2\sum_{1\le k<\ell\le n} c_{10}e^{-c_{11}\ell}.$$

Note that the sums on the right are uniformly bounded in n and that the constants c_{10}, c_{11} do not depend on η. Thus, we conclude that

$$\sup_{n,\eta}\left|\mathrm{Var}_\eta\left(\sum_{k=1}^n G_k\right) - \mathrm{Var}_\pi\left(\sum_{k=1}^n G_k\right)\right| < \infty. \tag{32}$$

Step 2. Note that

$$\frac{\nu_n}{n} = \frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}_\pi(G_k) + \frac{2}{n}\sum_{1\le k<\ell\le n}\mathrm{Cov}_\pi(G_k, G_\ell) = \mathrm{Var}_\pi(G_1) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}_\pi(G_1, G_{1+k}),$$

where in the last equality the change in the indices is due to the fact that π is the stationary distribution for the Markov chain $\{I_m\}_{m\ge 0}$. Therefore,

$$\mathrm{Var}_\pi(G_1) + 2\sum_{k=1}^{\infty}\mathrm{Cov}_\pi(G_1, G_{1+k}) - \frac{\nu_n}{n} = 2\sum_{k=n}^{\infty}\mathrm{Cov}_\pi(G_1, G_{1+k}) + 2\sum_{k=1}^{n-1}\frac{k}{n}\mathrm{Cov}_\pi(G_1, G_{1+k}).$$

It follows from (29) and (31) that

$$\left|\mathrm{Cov}_\pi(G_1, G_{1+k})\right| = \left|\pi\cdot(\mathbf{v}_k - \lambda\mathbf{g})\right| \le Ce^{-C'k}.$$

Therefore, for some $C'' > 0$

$$\left|\mathrm{Var}_\pi(G_1) + 2\sum_{k=1}^{\infty}\mathrm{Cov}_\pi(G_1, G_{1+k}) - \frac{\nu_n}{n}\right| \le 2\sum_{k=n}^{\infty}Ce^{-C'k} + \frac{2}{n}\sum_{k=1}^{n-1}Cke^{-C'k} \le \frac{C''}{n}.$$

Step 3. Since $E_i[G_1^2] = \sum_{k=1}^{\infty}(2k-1)P_i(G_1\ge k)$, recalling the formula in (22) for $P_i(G_1\ge k)$ we obtain

$$\begin{aligned}\mathbf{v}_0 &= \sum_{k=1}^{\infty}(2k-1)(D_pK)^k\mathbf{1} = (I+D_pK)(I-D_pK)^{-2}D_pK\mathbf{1}\\ &= (I-D_pK)^{-1}(I+D_pK)(I-D_pK)^{-1}\mathbf{p} = (I-D_pK)^{-1}(I+D_pK)\mathbf{g}.\end{aligned} \tag{33}$$

Using (33) and (29), we have that

$$\begin{aligned}\mathrm{Var}_\pi(G_1) &= \pi\big((I-D_pK)^{-1}(I+D_pK) - \lambda I\big)\mathbf{g} = \pi(I-D_pK)^{-1}\big((1-\lambda)I + (1+\lambda)D_pK\big)\mathbf{g}\\ &= (1-\lambda)\pi(I-D_pK)^{-1}\mathbf{g} + (1+\lambda)\pi D_pK(I-D_pK)^{-1}\mathbf{g}.\end{aligned}$$

Since $\pi = (1+\lambda)\mu(I-D_pK)$ this simplifies to

$$\mathrm{Var}_\pi(G_1) = \big(1-\lambda^2\big)\mu\mathbf{g} + (1+\lambda)^2\mu D_pK\mathbf{g}. \tag{34}$$

To compute $\mathrm{Cov}_\pi(G_1, G_{1+k})$ we need a formula for $\mathbf{v}_k$. Recall that $M_n(i,i') = P_i(G_1 = n, I_1 = i')$, so that by conditioning on G_1 and I_1 we obtain

$$E_i[G_1G_{1+k}] = \sum_{n=0}^{\infty}nM_n\big(i,i'\big)E_{i'}[G_k] = \sum_{n=0}^{\infty}ne_i\big(M_n\Pi^{k-1}\mathbf{g}\big).$$

Using this and the fact that $M_n = (D_pK)^nD_{1-p}K$ we get

$$\begin{aligned}\mathbf{v}_k &= \sum_{n=0}^{\infty}n\big(M_n\Pi^{k-1}\mathbf{g}\big) = \left(\sum_{n=0}^{\infty}n(D_pK)^n\right)D_{1-p}K\Pi^{k-1}\mathbf{g}\\ &= D_pK(I-D_pK)^{-2}D_{1-p}K\Pi^{k-1}\mathbf{g}.\end{aligned}$$

Now, note that if in this equation the matrix Π^{k-1} is replaced by $\mathbf{1}\pi$ (recall this is the matrix with all rows equal to π) then since $\pi\mathbf{g} = \lambda$ we have

$$\begin{aligned} D_pK(I-D_pK)^{-2}D_{1-p}K\mathbf{1}\pi\mathbf{g} &= \lambda D_pK(I-D_pK)^{-2}D_{1-p}K\mathbf{1} = \lambda D_pK(I-D_pK)^{-2}(\mathbf{1}-\mathbf{p}) \\ &= \lambda(I-D_pK)^{-1}D_pK(I-D_pK)^{-1}(\mathbf{1}-\mathbf{p}) = \lambda(I-D_pK)^{-1}D_pK\mathbf{1} \\ &= \lambda(I-D_pK)^{-1}\mathbf{p} = \lambda\mathbf{g}. \end{aligned}$$

Therefore, we can re-write the formula for $\mathrm{Cov}_\pi(G_1, G_{1+k})$ in (29) as

$$\mathrm{Cov}_\pi(G_1, G_{1+k}) = \pi \cdot (\mathbf{v}_k - \lambda\mathbf{g}) = \pi D_pK(I-D_pK)^{-2}D_{1-p}K\big(\Pi^{k-1} - \mathbf{1}\pi\big)\mathbf{g}.$$

Recalling the definition of $\mathbf{r}$ in (23), this implies that

$$\begin{aligned} \sum_{k=1}^{\infty}\mathrm{Cov}_\pi(G_1, G_{1+k}) &= \pi D_pK(I-D_pK)^{-2}D_{1-p}K\mathbf{r} \\ &= \pi(I-D_pK)^{-1}D_pK(I-D_pK)^{-1}D_{1-p}K\mathbf{r} = (1+\lambda)\mu D_pK\Pi\mathbf{r}, \end{aligned} \tag{35}$$

where the last equality follows from Lemma 3.2.

Combining (34) and (35), we obtain that

$$\nu = \big(1-\lambda^2\big)\mu\mathbf{g} + (1+\lambda)^2\mu D_pK\mathbf{g} + 2(1+\lambda)\mu D_pK\Pi\mathbf{r}.$$

To further simplify this, note that it follows from (24) and Corollary 3.5 that

$$\mathbf{g} = (I - \Pi + \mathbf{1}\pi)(\mathbf{r} + \lambda\mathbf{1}) = (I-\Pi)\mathbf{r} + \lambda\mathbf{1}.$$

Re-arranging this we get $\Pi\mathbf{r} = \mathbf{r} + \lambda\mathbf{1} - \mathbf{g}$, and putting this back into the above formula for ν we get

$$\begin{aligned} \nu &= \big(1-\lambda^2\big)\mu\mathbf{g} + (1+\lambda)^2\mu D_pK\mathbf{g} + 2(1+\lambda)\mu D_pK(\mathbf{r} + \lambda\mathbf{1} - \mathbf{g}) \\ &= \big(1-\lambda^2\big)\mu\mathbf{g} + \big(\lambda^2-1\big)\mu D_pK\mathbf{g} + 2(1+\lambda)\mu D_pK\mathbf{r} + 2(1+\lambda)\lambda\mu D_pK\mathbf{1} \\ &= \big(1-\lambda^2\big)\mu(I-D_pK)\mathbf{g} + 2(1+\lambda)\mu D_pK\mathbf{r} + 2(1+\lambda)\lambda\bar{p} \\ &= \big(1-\lambda^2\big)\mu\cdot\mathbf{p} + 2(1+\lambda)\mu D_pK\mathbf{r} + 2(1+\lambda)\lambda\bar{p} = (1+\lambda)\lambda + 2(1+\lambda)\mu D_pK\mathbf{r}, \end{aligned}$$

where in the second to last equality we used the formula (20) for $\mathbf{g}$, and in the last equality we used that $\mu\cdot\mathbf{p} = \bar{p} = \frac{\lambda}{1+\lambda}$. □

4. The backward BLP and parameter relationships

Throughout this section we will assume that we are in the critical case $\bar{p} = 1/2$. We shall discuss the backward BLP, give explicit formulas for parameters δ and $\tilde{\delta}$, and derive the key relationship between them.

Consider first the backward BLP V. To obtain the results about V from those for the forward BLP U:

- we need to replace $\mathbf{p}$ with $\mathbf{1}-\mathbf{p}$ everywhere in order to switch from counting "successes" to counting "failures";
- we have to account for the fact that V has one "immigrant" in each generation: recall that $V_i = F^i_{V_{i-1}+1}$ while $U_i = S^i_{U_{i-1}}$, $i \in \mathbb{N}$.

The above observations lead to the following statements whose proofs are identical to those for U. We shall state the results only for the critical case, since we use V solely in the critical setting.

Proposition 4.1. *Let $\bar p = 1/2$. For every distribution η on $\mathcal{R}$*

$$\lim_{n\to\infty}\frac{E_\eta[V_1|V_0=n]}{n}=1.$$

Moreover, there exist constants $c_{12}, c_{13} > 0$ such that for any $n \in \mathbb{N}$ and any distribution η on $\mathcal{R}$

$$\left|E_\eta[V_1|V_0=n]-n-(1+\eta\cdot\tilde{\mathbf{r}})\right|\le c_{12}e^{-c_{13}n},$$

where

$$\tilde{\mathbf{r}}=\mathbf{r}(\mathbf{1}-\mathbf{p},K)=(I-K+2\mathbf{p}\mu D_pK)^{-1}(\mathbf{1}-\mathbf{p})-\mathbf{1}.$$

Proposition 4.2. *Let $\bar p = 1/2$. There exists a constant $c_{14} > 0$ such that for any distribution η on $\mathcal{R}$*

$$\left|\frac{\operatorname{Var}_\eta(V_1|V_0=n)}{n}-\tilde\nu\right|\le\frac{c_{14}}{n},$$

where the constant $\tilde\nu$ is given by the formula

$$\tilde\nu=\nu(\mathbf{1}-\mathbf{p},K)=2+4\mu D_{1-p}K\tilde{\mathbf{r}}=4\mu D_{1-p}K(I-K+2\mathbf{p}\mu D_pK)^{-1}(\mathbf{1}-\mathbf{p}).\tag{36}$$

For reader's convenience we list the relevant parameters for U and V side by side.

$$\begin{aligned}&\pi=2\mu(1-D_pK)=2\mu D_{1-p}K; &&\tilde\pi=2\mu(1-D_{1-p}K)=2\mu D_pK;\\ &\mathbf{r}=\big(I-K+2(\mathbf{1}-\mathbf{p})\mu D_{1-p}K\big)^{-1}\mathbf{p}-\mathbf{1}; &&\tilde{\mathbf{r}}=(I-K+2\mathbf{p}\mu D_pK)^{-1}(\mathbf{1}-\mathbf{p})-\mathbf{1};\\ &\nu=2+4\mu D_pK\mathbf{r}=2+2\tilde\pi\mathbf{r}; &&\tilde\nu=2+4\mu D_{1-p}K\tilde{\mathbf{r}}=2+2\pi\tilde{\mathbf{r}}.\end{aligned}\tag{37}$$

In the above formulas, recall that:

- K is the transition matrix for the Markov chain R_j used to generate the cookie environment at each site. The row vector μ is the stationary distribution for this Markov chain.
- $\mathbf{p}$ is the column vector with ith entry $p(i)$, $\mathbf{1}$ is a column vector of all ones, and D_p and D_{1-p} are diagonal matrices with ith entry $p(i)$ and $1-p(i)$, respectively.
- The row vectors π and $\tilde\pi$ give the limiting distribution of the next "cookie" to be used after the nth failure or success, respectively, in the sequence of Bernoulli trials $\xi(j)$ at a site.
- The column vector $\mathbf{r}$ and the parameter ν are related to the asymptotics of the mean and variance of the forward BLP U as given in Propositions 3.1 and 3.6, and $\tilde{\mathbf{r}}$ and $\tilde\nu$ are related to the asymptotics of the backward BLP V in the same way by Propositions 4.1 and 4.2.

Now we can define the parameters δ and $\tilde\delta$ that appear in Theorems 1.6–1.8.

$$\delta=\frac{2\eta\cdot\mathbf{r}}{\nu},\qquad\tilde\delta=\frac{2\eta\cdot\tilde{\mathbf{r}}}{\tilde\nu}.\tag{38}$$

Note that the parameter δ can be computed in terms of $\mathbf{p}$, K and η. If we wish to make this dependence explicit we will write $\delta=\delta(\mathbf{p},K,\eta)$. In particular, with this notation we have that $\tilde\delta=\delta(\mathbf{1}-\mathbf{p},K,\eta)$.

We close this section with a justification of the statement of Theorem 1.6 that $\delta+\tilde\delta<1$.

Proposition 4.3. $\nu=\tilde\nu$ *and* $\delta+\tilde\delta=1-2/\nu$.

Proof. We shall need the following lemma.

Lemma 4.4. $\mathbf{r}+\tilde{\mathbf{r}}=(\mu(\mathbf{r}+\tilde{\mathbf{r}}))\mathbf{1}$. *In particular,* $\mathbf{r}+\tilde{\mathbf{r}}$ *is a constant multiple of* $\mathbf{1}$.

Let us postpone the proof of Lemma 4.4 and continue with the proof of Proposition 4.3. Recall that by Corollary 3.5 $\pi \cdot \mathbf{r} = 0$. Similarly, $\tilde{\pi} \cdot \tilde{\mathbf{r}} = 0$. Therefore, Lemma 4.4 and the formulas for ν and $\tilde{\nu}$ in (37) imply that

$$\tilde{\nu} = 2 + 2\pi(\mathbf{r} + \tilde{\mathbf{r}}) \overset{\text{L.4.4}}{=} 2 + 2\mu(\mathbf{r} + \tilde{\mathbf{r}}) \overset{\text{L.4.4}}{=} 2 + 2\tilde{\pi}(\mathbf{r} + \tilde{\mathbf{r}}) = \nu.$$

Moreover, from the above line we see that $\mu(\mathbf{r} + \tilde{\mathbf{r}}) = \nu/2 - 1$. Combining this with Lemma 4.4 we get that $\mathbf{r} + \tilde{\mathbf{r}} = (\nu/2 - 1)\mathbf{1}$. Then by (38) we have $\delta + \tilde{\delta} = 2\eta \cdot (\mathbf{r} + \tilde{\mathbf{r}})/\nu = 1 - 2/\nu$ as claimed. □

Proof of Lemma 4.4. Recall that $\mathbf{r}_n$ is the column vector with ith entry

$$r_n(i) = E_i^{U,n}[U_1] - n = E_i[S_n] - n,$$

where S_n is the number of successes before the nth failure in the sequence of Bernoulli trials $\{\xi(j)\}_{j\geq 1}$. Let $Z_n = \sum_{j=1}^n \xi(j)$ be the number of successes in the first n Bernoulli trials. If $Z_n = k$, then there are $n - k$ failures among the first n Bernoulli trials and so S_n is equal to k plus the number of successes before the kth failure in shifted Bernoulli sequence $\{\xi(j)\}_{j\geq n+1}$. Therefore, by conditioning on Z_n and R_{n+1} (the type of the $(n+1)$st cookie), we obtain that

$$\begin{aligned}
r_n(i) &= \sum_{i'\in\mathcal{R}} \sum_{k=0}^{n} P_i\bigl(Z_n = k, R_{n+1} = i'\bigr)\bigl(k + E_{i'}[S_k] - n\bigr) \\
&= \sum_{i'\in\mathcal{R}} \sum_{k=0}^{n} P_i\bigl(Z_n = k, R_{n+1} = i'\bigr)\bigl(2k + r_k\bigl(i'\bigr) - n\bigr) \\
&= E_i[2Z_n - n] + \sum_{i'\in\mathcal{R}} \sum_{k=0}^{n} P_i\bigl(Z_n = k, R_{n+1} = i'\bigr) r_k\bigl(i'\bigr) \\
&= \sum_{j=1}^{n} \mathbb{E}_i\bigl[2p(R_j) - 1\bigr] + e_i K^n \mathbf{r} + \sum_{i'\in\mathcal{R}} \sum_{k=0}^{n} P_i\bigl(Z_n = k, R_{n+1} = i'\bigr)\bigl(r_k\bigl(i'\bigr) - r\bigl(i'\bigr)\bigr).
\end{aligned}$$

Since $\|\mathbf{r}_k - \mathbf{r}\|_\infty$ decreases exponentially in k and $\lim_{n\to\infty} P_i(Z_n = k, R_{n+1} = i') = 0$ for every $k \geq 0$ and $i' \in \mathcal{R}$, it follows from the dominated convergence theorem that the last sum vanishes as $n \to \infty$. Since $\mathbf{r}_n \to \mathbf{r}$ as $n \to \infty$, we have shown that

$$r(i) = \lim_{n\to\infty} \left\{ \sum_{j=1}^{n} \mathbb{E}_i\bigl[2p(R_j) - 1\bigr] + e_i K^n \mathbf{r} \right\}, \quad \forall i \in \mathcal{R}. \tag{39}$$

where the limit on the right necessarily converges. Similarly, a symmetric argument interchanging the roles of failures and successes yields

$$\tilde{r}(i) = \lim_{n\to\infty} \left\{ \sum_{j=1}^{n} \mathbb{E}_i\bigl[1 - 2p(R_j)\bigr] + e_i K^n \tilde{\mathbf{r}} \right\}, \quad \forall i \in \mathcal{R}. \tag{40}$$

Since clearly $Z_n + \tilde{Z}_n = n$ for all n, it follows from (39) and (40) that

$$\mathbf{r} + \tilde{\mathbf{r}} = \lim_{n\to\infty} K^n(\mathbf{r} + \tilde{\mathbf{r}}) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} K^j(\mathbf{r} + \tilde{\mathbf{r}}) = (\mathbf{1}\mu)(\mathbf{r} + \tilde{\boldsymbol{r}}).$$

(Note that the second limit above is needed since we did not assume that the Markov chain R_j is aperiodic on the closed irreducible subset $\mathcal{R}_0$.) □

5. The non-critical case

In this section we consider the non-critical case $\bar{p} = \mu \cdot \mathbf{p} \neq 1/2$ and give the proof of Theorem 1.5. The proof will be obtained through an analysis of the forward and backward BLP from Section 2. Theorems 2.6 and 2.7, which were the basis for the proofs of Theorems 1.6–1.8, cannot be directly applied in the non-critical case $\bar{p} \neq 1/2$. The main idea in this section will be to construct a coupling of the non-critical forward/backward BLP with a corresponding critical forward/backward BLP and then use this coupling and the connection of the BLP with the ERW to obtain the conclusions of Theorem 1.5.

5.1. *Coupling*

Let K be a transition probability matrix for a Markov chain on $\mathcal{R} = \{1, 2, \dots, N\}$ with some initial distribution η. We will assume that this Markov chain satisfies the assumptions stated at the beginning of Section 1.1. In particular, it has a unique stationary distribution μ supported on a closed irreducible set $\mathcal{R}_0 \subseteq \mathcal{R}$. Suppose that we are given two functions $p_0, p_1 : \mathcal{R} \to (0, 1)$ such that $p_1 \geq p_0$. Let $\mathbf{p_i} := (p_i(1), p_i(2), \dots, p_i(N))$, $\mathbf{i} = 0, 1$, and assume that $\mu \cdot \mathbf{p_1} > \mu \cdot \mathbf{p_0} = 1/2$.

Next, we expand the state space for the Markov chain to be $\hat{\mathcal{R}} = \{1, 2, \dots, 2N\}$, and for any $\varepsilon \in (0, 1]$ we let $\hat{K}_\varepsilon$ be a $2N \times 2N$ transition matrix and $\hat{\mathbf{p}}$ be a column vector of length $2N$ given by

$$\hat{K}_\varepsilon = \left(\begin{array}{c|c} (1-\varepsilon)K & \varepsilon K \\ \hline O_N & K \end{array}\right) \quad \text{and} \quad \hat{\mathbf{p}} = \begin{pmatrix} \mathbf{p_1} \\ \mathbf{p_0} \end{pmatrix}, \tag{41}$$

where O_N denotes an $N \times N$ matrix of all zeros. It is clear that the unique stationary distribution for $\hat{K}_\varepsilon$ is given by $\hat{\mu} = (\mathbf{0}, \mu)$, where $\mathbf{0}$ is a row vector of N zeros and μ is the stationary distribution for the transition matrix K. Moreover, $\hat{\mu} \cdot \hat{\mathbf{p}} = \mu \cdot \mathbf{p_0} = 1/2$ so that $\hat{K}_\varepsilon$ and $\hat{\mathbf{p}}$ correspond to a critical ERW.

Let $\hat{U}^\varepsilon$ and $\hat{V}^\varepsilon$ be the associated forward and backward BLP for the critical ERW corresponding to $\hat{\mathbf{p}}$, $\hat{K}_\varepsilon$, and with initial distribution $\hat{\eta}_\varepsilon = ((1-\varepsilon)\eta, \varepsilon\eta)$. Let U and V be the forward and backward BLP for the non-critical case corresponding to $\mathbf{p_1}$, K, and the initial distribution η. We claim that for any $\varepsilon \in (0, 1]$ these BLPs can be coupled so that:

- $U_k \geq \hat{U}^\varepsilon_k$ for all $k \geq 0$;
- $V_k \leq \hat{V}^\varepsilon_k$ for all $k \geq 0$.

To this end, recall that $\mathbb{P}_\eta$ is the distribution for the i.i.d. family of Markov chains $\{R^x_j\}_{j \geq 0}$, $x \in \mathbb{Z}$, with transition matrix K and initial condition η. We can enlarge the probability space so that the measure $\mathbb{P}_\eta$ also includes an i.i.d. sequence $\{\gamma^x_\varepsilon\}_{x \in \mathbb{Z}}$ of Geometric(ε) random variables (that is $\mathbb{P}_\eta(\gamma^x_\varepsilon = k) = (1-\varepsilon)^k \varepsilon$ for $k \geq 0$) that is also independent of the family of Markov chains $\{R^x_j\}_{j \geq 1}$, $x \in \mathbb{Z}$. We let $\omega_x(j) = p_1(R^x_j)$ and also define a different cookie environment $\hat{\omega}^\varepsilon = \{\hat{\omega}^\varepsilon_x(j)\}_{x \in \mathbb{Z}, j \geq 1}$ by

$$\hat{\omega}^\varepsilon_x(j) = \begin{cases} p_1(R^x_j), & j \leq \gamma^x_\varepsilon, \\ p_0(R^x_j), & j > \gamma^x_\varepsilon. \end{cases}$$

It is clear from this construction that the cookie environments ω and $\hat{\omega}$ are coupled so that $\hat{\omega}^\varepsilon_x(j) \leq \omega_x(j)$ for all $x \in \mathbb{Z}$ and $j \geq 1$. Given such a pair $(\omega, \hat{\omega}^\varepsilon)$ of coupled cookie environments, let $\{\xi_x(j)\}_{x \in \mathbb{Z}, j \geq 1}$ and $\{\hat{\xi}^\varepsilon_x(j)\}_{x \in \mathbb{Z}, j \geq 1}$ be families of independent Bernoulli random variables with $\xi_x(j) \sim \text{Bernoulli}(\omega_x(j))$ and $\hat{\xi}^\varepsilon_x(j) \sim \text{Bernoulli}(\hat{\omega}^\varepsilon_x(j))$ that are coupled to have $\hat{\xi}^\varepsilon_x(j) \leq \xi_x(j)$ for all $x \in \mathbb{Z}$ and $j \geq 1$. If U and V are the forward and backward BLP constructed from the Bernoulli family $\{\xi_x(j)\}_{x,j}$ and $\hat{U}^\varepsilon$ and $\hat{V}^\varepsilon$ are the forward and backward BLP constructed from the Bernoulli family $\{\hat{\xi}^\varepsilon_x(j)\}_{x,j}$, then the couplings $U_k \geq \hat{U}^\varepsilon_k$ and $V_k \leq \hat{V}^\varepsilon_k$ for all $k \geq 0$ follow immediately. Moreover, it is easy to see that U, V, $\hat{U}^\varepsilon$ and $\hat{V}^\varepsilon$ have the required marginal distributions under this coupling.

As noted above, the forward and backward BLP $\hat{U}^\varepsilon$ and $\hat{V}^\varepsilon$ are critical BLP to which Theorems 2.6 and 2.7 apply. In the applications of these theorems, however, one must replace the parameter $\delta = \delta(\mathbf{p}, K, \eta)$ by

$$\hat{\delta}_\varepsilon = \hat{\delta}_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta) := \delta(\hat{\mathbf{p}}, \hat{K}_\varepsilon, \hat{\eta}_\varepsilon).$$

The following lemma shows, in particular, that the parameter $\hat{\delta}_\varepsilon$ can be made arbitrarily large by letting $\varepsilon \to 0$.

Lemma 5.1. *Let $\varepsilon \in (0,1]$, η be an arbitrary distribution on $\mathcal{R}$, and $\hat{U}^\varepsilon$ and $\hat{V}^\varepsilon$ be the BLPs constructed above. Then*

$$\hat{\delta}_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta) = \delta(\mathbf{p_0}, K, \eta) + \frac{4\eta \cdot (\sum_{j=1}^\infty (1-\varepsilon)^j K^{j-1}(\mathbf{p_1} - \mathbf{p_0}))}{\nu(\mathbf{p_0}, K)}. \tag{42}$$

In particular, $\hat{\delta}_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta)$ is a continuous strictly decreasing function of ε on $(0,1]$ which equals $\delta(\mathbf{p_0}, K, \eta)$ when $\varepsilon = 1$ and satisfies

$$\lim_{\varepsilon \downarrow 0} \hat{\delta}_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta) = \infty. \tag{43}$$

Proof. It follows from the formula for δ given in (38) that

$$\hat{\delta}_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta) = \frac{2((1-\varepsilon)\eta, \varepsilon\eta) \cdot \hat{\mathbf{r}}_\varepsilon}{\hat{\nu}_\varepsilon}, \quad \text{where } \hat{\mathbf{r}}_\varepsilon = \mathbf{r}(\hat{\mathbf{p}}, \hat{K}_\varepsilon) \text{ and } \hat{\nu}_\varepsilon = \nu(\hat{\mathbf{p}}, \hat{K}_\varepsilon).$$

Therefore, if $\hat{\mathbf{r}}_\varepsilon = (\hat{r}_\varepsilon(1), \hat{r}_\varepsilon(2), \dots, \hat{r}_\varepsilon(2N))^t$ then (42) will follow of we show that for all $\varepsilon \in (0,1]$

$$\hat{\nu}_\varepsilon = \nu(\hat{\mathbf{p}}, \hat{K}_\varepsilon) = \nu(\mathbf{p_0}, K), \quad \text{and} \tag{44}$$

$$\hat{\mathbf{r}}_\varepsilon = \hat{\mathbf{r}}(\hat{\mathbf{p}}, \hat{K}_\varepsilon) = \left(\mathbf{r}(\mathbf{p_0}, K) + 2\sum_{j=1}^\infty (1-\varepsilon)^j K^{j-1}(\mathbf{p_1} - \mathbf{p_0}), \mathbf{r}(\mathbf{p_0}, K) \right). \tag{45}$$

Note that (45) states, in part, that the last N coordinates of $\hat{\mathbf{r}}_\varepsilon$ are equal to $\mathbf{r}(\mathbf{p_0}, K)$ for any $\varepsilon \in (0,1]$. This follows easily from the fact that the bottom right $N \times N$ block of the transition matrix $\hat{K}_\varepsilon$ in (41) is equal to K. Similarly, since the states $i = 1, 2, \dots, N$ are transient for the Markov chain with transition matrix $\hat{K}_\varepsilon$, it follows that the stationary distribution $\tilde{\pi}$ corresponding to the pair $(\hat{\mathbf{p}}, \hat{K}_\varepsilon)$ is given by $\tilde{\pi}(\hat{\mathbf{p}}, \hat{K}_\varepsilon) = (\mathbf{0}, \tilde{\pi}(\mathbf{p_0}, K))$. Thus, the formula for the parameter ν in (37) implies that

$$\hat{\nu}_\varepsilon = 2 + 2\tilde{\pi}(\hat{\mathbf{p}}, \hat{K}_\varepsilon) \cdot \hat{\mathbf{r}}_\varepsilon = 2 + 2\tilde{\pi}(\mathbf{p_0}, K) \cdot \mathbf{r}(\mathbf{p_0}, K) = \nu(\mathbf{p_0}, K).$$

This proves (44), and it remains to show (45) for the first N coordinates of $\hat{\mathbf{r}}_\varepsilon$.

Applying the formula for $r(i)$ from (39), we have that

$$\begin{aligned} \hat{r}_\varepsilon(i) &= \lim_{n\to\infty} \left\{ \sum_{j=1}^n \mathbb{E}_i\left[2\left(p_1(R_j)\mathbf{1}_{\{j \le \gamma_\varepsilon\}} + p_0(R_j)\mathbf{1}_{\{j > \gamma_\varepsilon\}}\right) - 1\right] + e_i \hat{K}_\varepsilon^n \hat{\mathbf{r}}_\varepsilon \right\} \\ &= \lim_{n\to\infty} \left\{ \sum_{j=1}^n 2(1-\varepsilon)^j e_i K^{j-1}(\mathbf{p}_1 - \mathbf{p}_0) + \sum_{j=1}^n \mathbb{E}_i\left[2p_0(R_j) - 1\right] + e_i \hat{K}_\varepsilon^n \hat{\mathbf{r}}_\varepsilon \right\}. \end{aligned} \tag{46}$$

It follows from the form of the matrix $\hat{K}_\varepsilon$ and the fact that the last N entries of the vector $\hat{r}_\varepsilon$ are equal to $\mathbf{r}(\mathbf{p_0}, K)$ that

$$\lim_{n\to\infty} \left| e_i \hat{K}_\varepsilon^n \hat{\mathbf{r}}_\varepsilon - e_i K^n \mathbf{r}(\mathbf{p_0}, K) \right| = 0.$$

Together with (39) this implies that the limit of the last two terms in (46) is the ith entry of the vector $\mathbf{r}(\mathbf{p_0}, K)$. Since the limit of the first term in (46) is the corresponding infinite sum we obtain (45).

Finally, to prove (43), it is enough to show that

$$\sum_{j=0}^\infty \eta K^j(\mathbf{p}_0 - \mathbf{p}_1) = \infty. \tag{47}$$

Recalling that $p_1 \geq p_0$ and $\mu \cdot \mathbf{p_1} > \mu \cdot \mathbf{p_0} = 1/2$ we conclude that there is a strict inequality $p_1(i) > p_0(i)$ for some state $i \in \mathcal{R}_0$ which is therefore a recurrent state for the Markov chain R_j. Since the ith entry of $\sum_{j=0}^{\infty} \eta K^j$ equals the expected number of times the Markov chain R_j visits the state i, (47) follows. □

5.2. *Proof of Theorem* 1.5

Without loss of generality we may assume that $\bar{p} > 1/2$. We shall apply Lemma 5.1. Choose $\mathbf{p_1} = \mathbf{p}$. Since $\bar{p} > 1/2$, there exists $\mathbf{p_0} = (p_0(1), p_0(2), \dots, p_0(N))^t \in (0,1)^N$ such that:

- $p_0(i) \leq p_1(i)$ for all $i \in \mathcal{R}$.
- $\mu \cdot \mathbf{p_0} = 1/2$.

Let the distribution η on $\mathcal{R}$ be fixed. By Lemma 5.1 we may choose an $\varepsilon \in (0,1]$ so that $\hat{\delta}_\varepsilon = \delta_\varepsilon(\mathbf{p_1}, \mathbf{p_0}, K, \eta) > 4$, and we will keep this choice of ε fixed for the remainder of the proof.

Transience: Since $\hat{\delta}_\varepsilon > 1$, it follows from Theorem 2.6 that $\sigma_0^{\hat{U}^\varepsilon} = \infty$ with positive probability for any initial condition $\hat{U}_0^\varepsilon = y \geq 1$ for the forward BLP $\hat{U}^\varepsilon$. Since the above coupling of the forward BLP is such that $U_k \geq \hat{U}_k^\varepsilon$ for all $k \geq 1$, this implies that $P_\eta^{U,y}(\sigma_0^U = \infty) > 0$ for any $y \geq 1$. From this, the proof that the ERW is transient to the right is the same as in the proof of Theorem 1.6.

Ballisticity and Gaussian limits: Since $\hat{\delta}_\varepsilon > 4$, it follows from Theorem 2.7 that $\sigma_0^{\hat{V}^\varepsilon}$ and $\sum_{i=0}^{\sigma_0^{\hat{V}^\varepsilon}-1} \hat{V}_i^\varepsilon$ have finite second moments when the backward BLP $\hat{V}^\varepsilon$ is started with $\hat{V}_0^\varepsilon = 0$. Since the above coupling is such that $V_k \leq \hat{V}_k^\varepsilon$ for all $k \geq 0$ this implies that

$$E_\eta^{V,0}\big[\big(\sigma_0^V\big)^2\big] < \infty \quad \text{and} \quad E_\eta^{V,0}\left[\left(\sum_{i=0}^{\sigma_0^V-1} V_i\right)^2\right] < \infty. \tag{48}$$

As noted in the proof of Theorem 1.7, a formula for the limiting speed v_0 when the ERW is transient can be given by

$$\frac{1}{v_0} = 1 + 2\frac{E_\eta^{V,0}[\sum_{i=0}^{\sigma_0^V-1} V_i]}{E_\eta^{V,0}[\sigma_0^V]}.$$

Since (48) implies that the fraction on the right is finite, we can conclude that the limiting speed $v_0 > 0$. The proof of the Gaussian limiting distributions for the ERW is the same as for the proof of the limiting distributions of the critical ERWs in the case $\delta > 4$ since all that is needed are the finite second moments for the backward BLP given in (48).

6. Proofs of Theorems 2.6 and 2.7

We shall discuss the proof of Theorem 2.7, since Theorem 2.6 can be derived in exactly the same way. The main idea behind the proof of Theorem 2.7 is very simple. The rescaled critical backward BLP can be approximated (see Lemma 6.1 below) by a constant multiple of a squared Bessel process of the generalized dimension

$$\frac{4(1+\eta\cdot\tilde{\mathbf{r}})}{\tilde{\nu}} \overset{(38),\,\text{L. }4.3}{=} \frac{4}{\nu} + 2\tilde{\delta} \overset{\text{L. }4.3}{=} 2(1-\delta). \tag{49}$$

The squared Bessel process of dimension 2 is just the square of the distance of the standard planar Brownian motion from the origin, and this dimension is critical. For dimensions less than 2 the squared Bessel process hits 0 with probability 1. For dimensions 2 or higher the origin is not attainable. Thus, $\delta = 0$ should be a critical value for V. This suggests that the process dies out with probability 1 if $\delta > 0$ and has a positive probability of survival if $\delta < 0$. Moreover, the tail decay exponents for $\delta > 0$ can also be read off from those of the corresponding squared Bessel process. The boundary case $\delta = 0$ is delicate and has to be handled separately.

Nevertheless, a lot of work needs to be done to turn the above idea into a proof. This is accomplished in [23] and [9] for the model with bounded cookie stacks. The proofs of the respective results in the cited above papers can be repeated *verbatim* provided that we reprove several lemmas which depend on the specifics of the backward BLP V. Therefore, below we only provide a rough sketch of the proof and state several lemmas which depend on the properties of our V. These lemmas cover all $\delta \in \mathbb{R}$ and are used later in the section to discuss the three cases: $\delta > 0$, $\delta = 0$, and $\delta < 0$. Their proofs are given in the next section.

6.1. *Sketch of the proof*

Let us start with the already mentioned approximation by the squared Bessel process.[8]

Lemma 6.1 (Diffusion approximation, Lemma 3.1 of [23], Lemma 3.4 of [9]). *Fix an arbitrary $\varepsilon > 0$, $y > \varepsilon$, and a sequence $y_n \to y$ as $n \to \infty$. Define $Y^{\varepsilon,n}(t) = \frac{V_{\lfloor nt \rfloor \wedge \sigma_{\varepsilon n}}}{n}$, $t \geq 0$. Then, under P_η^{V,ny_n} the process $Y^{\varepsilon,n}$ converges in the Skorokhod (J_1) topology to $Y(\cdot \wedge \sigma_\varepsilon^Y)$ where Y is the solution of*

$$dY(t) = (1 + \eta \cdot \tilde{\mathbf{r}})\, dt + \sqrt{\nu Y(t)}\, dW(t), \quad Y(0) = y. \tag{50}$$

Remark. *For the ERW with bounded cookie stacks the convergence is known to hold up to σ_0^Y (or in $D([0,\infty))$ if $\sigma_0^Y = \infty$) as long as the corresponding squared Bessel process has dimension other than 2 (see [26, Theorem 3.4] for the forward branching process). But such result does not seem to significantly shorten the proof of Theorem 2.7, and we shall not show it.*

To prove (7) we need some tools to handle V when it starts with $y \ll \varepsilon n$ or falls below εn. The idea again comes from the properties of Y. It is easy to check that Y^δ for $\delta \neq 0$ (and $\ln Y$ for $\delta = 0$) is a local martingale. Then by a standard calculation we get that for all $a > 1$ and $j \in \mathbb{N}$

$$P\big(\tau_{a^{j-1}}^Y < \tau_{a^{j+1}}^Y \,|\, Y(0) = a^j\big) = \frac{a^\delta}{1 + a^\delta}.$$

The same power (logarithm for $\delta = 0$) of the rescaled process V is close to a martingale, and we can prove a similar result for all $\delta \in \mathbb{R}$ (for $\delta = 0$ it is sufficient to set $a = 2$ below).

Lemma 6.2 (Exit distribution, Lemma 5.2 of [23], Lemma A.3 of [9]). *Let $a > 1$, $|x - a^j| \leq a^{2j/3}$, and γ be the exit time from (a^{j-1}, a^{j+1}). Then for all sufficiently large $j \in \mathbb{N}$*

$$\left| P_\eta^{V,x}\big(V_\gamma \leq a^{j-1}\big) - \frac{a^\delta}{a^\delta + 1} \right| \leq a^{-j/4}.$$

One of the estimates needed for the proof of Lemma 6.2 is the following lemma which shows that the process V exits the interval (a^{j-1}, a^j) not too far below a^{j-1} or above a^j.

Lemma 6.3 ("Overshoot," Lemma 5.1 of [23]). *There are constants $c_{15}, c_{16} > 0$ and $N \in \mathbb{N}$ such that for all $x \geq N$, $y \geq 0$, and every initial distribution η of the environment Markov chain*

$$\max_{0 \leq z < x} P_\eta^{V,z}(V_{\tau_x} > x + y \,|\, \tau_x < \sigma_0) \leq c_{15}\big(e^{-c_{16} y^2/x} + e^{-c_{16} y}\big)$$

and

$$\max_{x < z < 4x} P_\eta^{V,z}(V_{\sigma_x \wedge \tau_{4x}} < x - y) \leq c_{15} e^{-c_{16} y^2/x}.$$

[8] Following the number of each lemma in this subsection are the corresponding results in the literature which it replaces.

Lemma 6.3 shows that the process V typically exits the interval (a^{j-1}, a^j) close enough to the boundary so that we can repeatedly apply Lemma 6.2 to couple V with a birth-and-death-like process to obtain the following estimate on exit probabilities from large intervals (a^ℓ, a^u) and ultimately handle V when it is below εn.

Lemma 6.4 (Lemma 5.3 of [23] and Lemma A.1 of [9]). *For each $a \in (1,2]$ there is an $\ell_0 \in \mathbb{N}$ and a small positive number λ such that if $\ell, m, u, x \in \mathbb{N}$ satisfy $\ell_0 \le \ell < m < u$ and $|x - a^m| < a^{2m/3}$ then*

$$\frac{h^-(m) - h^-(\ell)}{h^-(u) - h^-(\ell)} \le P_\eta^{V,x}\big(\sigma_{a^\ell}^V > \tau_{a^u}^V\big) \le \frac{h^+(m) - h^+(\ell)}{h^+(u) - h^+(\ell)},$$

where for $j \ge 1$

$$h^\pm(j) = \begin{cases} \prod_{i=1}^j (a^\delta \mp a^{-\lambda i}), & \text{if } \delta > 0; \\ \prod_{i=1}^j (a^{-\delta} \mp a^{-\lambda i})^{-1}, & \text{if } \delta < 0; \\ j \mp \frac{1}{j}, & \text{if } \delta = 0. \end{cases}$$

Moreover, for $\delta \ne 0$ there are $K_i : \mathbb{N} \to (0,\infty)$, $i = 1,2$, such that $K_i(\ell) \to 1$ as $\ell \to \infty$ and for all $j > \ell$

$$K_1(\ell) a^{(j-\ell)\delta} \le \frac{h^\pm(j)}{h^\pm(\ell)} \le K_2(\ell) a^{(j-\ell)\delta}. \tag{51}$$

Proof. The proof is essentially identical to the proof of [23, Lemma 5.3] and relies only on the properties proved already in Lemmas 6.2, 6.3 and the fact that the BLP are naturally monotone with respect to their initial value V_0. The proof of [23, Lemma 5.3] was given for $\delta > 0$, but essentially the same proof holds for $\delta \le 0$ with only minor changes needed due to the form of the functions $h^\pm(j)$. □

The proofs of Lemmas 6.1–6.3 will be given in Section 7. We close this section with a brief discussion of how to finish the proof of Theorem 2.7 using these lemmas in the cases $\delta > 0$, $\delta = 0$ and $\delta < 0$.

6.2. *Case $\delta > 0$*

Substituting the lemmas from the above subsection for the corresponding lemmas in [23] and repeating the proof given in this paper we obtain (7) for $\delta > 0$, which, in particular, implies that $P_\eta^{V,y}(\sigma_0^V < \infty) = 1$ for all $\delta > 0$.

6.3. *Case $\delta = 0$*

The proof of (7) for $\delta = 0$ is identical to the proof of [9, Theorem 2.1]. All the ingredients which depend on the specifics of V are contained in the lemmas above. Note again that while the 2-dimensional squared Bessel process never hits zero, (7) implies that $P_\eta^{V,y}(\sigma_0^V < \infty) = 1$ for $\delta = 0$.

6.4. *Case $\delta < 0$*

All we need to show is that $P_\eta^{V,y}(\sigma_0^V = \infty) > 0$ for every $y \ge 0$. Notice that by the strong Markov property and monotonicity of the BLP with respect to the starting point for every $m > \ell > \max\{\ell_0, \log_2(y+1)\}$

$$\begin{aligned} P_\eta^{V,y}\big(\sigma_0^V = \infty\big) &\ge P_\eta^{V,y}\big(\sigma_0^V = \infty, \tau_{2^m}^V < \sigma_0^V\big) \ge P_\eta^{V,2^m}\big(\sigma_0^V = \infty\big) P_\eta^{V,y}\big(\tau_{2^m}^V < \sigma_0^V\big) \\ &\ge P_\eta^{V,2^m}\big(\sigma_{2^\ell}^V = \infty\big) P_\eta^{V,y}\big(\tau_{2^m}^V < \sigma_0^V\big). \end{aligned}$$

By the lower bound of Lemma 6.4 and (51) we can choose and fix sufficiently large $m > \ell$ so that

$$P_\eta^{V,2^m}\big(\sigma_{2^\ell}^V = \infty\big) = \lim_{u\to\infty} P_\eta^{V,2^m}\big(\sigma_{2^\ell}^V > \tau_{2^u}^V\big) \ge 1 - \frac{h^-(m)}{h^-(\ell)} > 0.$$

Moreover, by the ellipticity of the environment, $P_\eta^{V,y}(\tau_{2^m}^V < \sigma_0^V) > 0$, and we conclude that $P_\eta^{V,y}(\sigma_0^V = \infty) > 0$ for every $y \ge 0$.

7. Proofs of Lemmas 6.1–6.3

We shall prove a more general diffusion approximation result and then derive from it Lemma 6.1.

Lemma 7.1 (Abstract lemma). *Let $b \in \mathbb{R}$, $D > 0$, and $Y(t)$, $t \ge 0$, be a solution of*

$$dY(t) = b\,dt + \sqrt{DY(t)^+}\,dW(t), \quad Y(0) = x > 0, \tag{52}$$

where $W(t)$, $t \ge 0$, is the standard Brownian motion.[9] *Let integer-valued Markov chains $Z_n := (Z_{n,k})_{k\ge 0}$ satisfy the following conditions*:

(i) *there is a sequence $N_n \in \mathbb{N}$, $N_n \to \infty$, $N_n = o(n)$ as $n \to \infty$, $f : \mathbb{N} \to [0,\infty)$, $f(x) \to 0$ as $x \to \infty$, and $g : \mathbb{N} \to [0,\infty)$, $g(x) \searrow 0$ as $x \to \infty$ such that*

(E) $\left|E(Z_{n,1} - Z_{n,0} | Z_{n,0} = m) - b\right| \le f(m \vee N_n)$;

(V) $\left|\dfrac{\mathrm{Var}(Z_{n,1} | Z_{n,0} = m)}{m \vee N_n} - D\right| \le g(m \vee N_n)$;

(ii) *for each $T, r > 0$*

$$E\Big(\max_{1\le k\le (Tn)\wedge \tau_{rn}^{Z_n}} (Z_{n,k} - Z_{n,k-1})^2\Big) = o\big(n^2\big) \quad \text{as } n \to \infty,$$

where $\tau_x^{Z_n} = \inf\{k \ge 0 : Z_{n,k} \ge x\}$.

Set $Z_{n,0} = \lfloor nx_n \rfloor$, $x_n \to x$ as $n \to \infty$, and $Y_n(t) = Z_{n,\lfloor nt\rfloor}/n$, $t \ge 0$. Then $Y_n \overset{J_1}{\Rightarrow} Y$ as $n \to \infty$, where $\overset{J_1}{\Rightarrow}$ denotes convergence in distribution with respect to the Skorokhod (J_1) topology.

Proof. The proof is based on [12, Theorem 4.1, p. 354]. We need to check the well-posedness of the martingale problem for

$$A = \left\{\left(f, Gf = \frac{D}{2}x^+ \frac{\partial^2}{\partial x^2} + b\frac{\partial}{\partial x}\right) : f \in C_c^\infty(\mathbb{R})\right\}$$

on $C_{\mathbb{R}}[0,\infty)$ and conditions (4.1)–(4.7) of that theorem. The well-posedness of the martingale problem follows from [12, Corollary 3.4, p. 295] and the fact that the existence and distributional uniqueness hold for solutions of (52) with arbitrary initial distributions.[10] Thus we only need to check the conditions of the theorem in [12].

To this end, define processes $A_n(t)$ and $B_n(t)$ by

$$A_n(t) := \frac{1}{n^2}\sum_{k=1}^{\lfloor nt\rfloor} \mathrm{Var}(Z_{n,k} | Z_{n,k-1}); \qquad B_n(t) := \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor} E(Z_{n,k} - Z_{n,k-1} | Z_{n,k-1}).$$

It is elementary to check that the processes $M_n(t) := Y_n(t) - B_n(t)$ and $M_n^2(t) - A_n(t)$, $t \ge 0$, are martingales for all $n \in \mathbb{N}$. Therefore, it is sufficient to check the following five conditions for any $T, r > 0$.

$$\lim_{n\to\infty} E\Big[\sup_{t\le T\wedge \tau_r^{Y_n}} \big|Y_n(t) - Y_n(t-)\big|^2\Big] = 0, \tag{53}$$

$$\lim_{n\to\infty} E\Big[\sup_{t\le T\wedge \tau_r^{Y_n}} \big|B_n(t) - B_n(t-)\big|^2\Big] = 0, \tag{54}$$

---

[9] Let us remark that $X(t) := 4Y(t)/D$ satisfies $dX(t) = (4b/D)\,dt + 2\sqrt{X(t)^+}\,dB(t)$ and, thus, is a squared Bessel process of the generalized dimension $4b/D$.

[10] A more detailed discussion of (52) can be found immediately following (3.1) in [26].

$$\lim_{n\to\infty} E\Big[\sup_{t\le T\wedge\tau_r^{Y_n}} \big|A_n(t)-A_n(t-)\big|\Big]=0, \tag{55}$$

$$\lim_{n\to\infty} \sup_{t\le T\wedge\tau_r^{Y_n}} \big|B_n(t)-bt\big|=0, \quad P\text{-a.s.}, \tag{56}$$

$$\lim_{n\to\infty} \sup_{t\le T\wedge\tau_r^{Y_n}} \left|A_n(t)-D\int_0^t Y_n^+(s)\,ds\right|=0, \quad P\text{-a.s.} \tag{57}$$

Note, in the above conditions that $\tau_r^{Y_n}=\inf\{t\ge 0: Y_n(t)\ge r\}=\tau_{rn}^{Z_n}/n$.

Condition (53) is simply a restatement of condition (ii) in the statement of Lemma 6.1. Conditions (54) and (55) follow from conditions (i)(E) and (i)(V), respectively. Indeed,

$$\begin{aligned}\lim_{n\to\infty} E\Big[\sup_{t\le T\wedge\tau_r^{Y_n}} \big|B_n(t)-B_n(t-)\big|^2\Big] &= \lim_{n\to\infty}\frac{1}{n^2}E\Big[\max_{1\le k\le (Tn)\wedge\tau_{rn}^{Z_n}} \big(E[Z_{n,k}-Z_{n,k-1}|Z_{n,k-1}]\big)^2\Big]\\ &\overset{\text{(i)(E)}}{\le} \lim_{n\to\infty}\frac{(b+\|f\|_\infty)^2}{n^2}=0,\end{aligned}$$

and similarly

$$\begin{aligned}\lim_{n\to\infty} E\Big[\sup_{t\le T\wedge\tau_r^{Y_n}} \big|A_n(t)-A_n(t-)\big|\Big] &= \lim_{n\to\infty}\frac{1}{n^2}E\Big[\max_{1\le k\le (Tn)\vee\tau_{rn}^{Z_n}} \operatorname{Var}(Z_{n,k}|Z_{n,k-1})\Big]\\ &\overset{\text{(i)(V)}}{\le} \lim_{n\to\infty}\frac{(nr)(D+\|g\|_\infty)}{n^2}=0,\end{aligned}$$

where in the last inequality we used that $Z_{n,k-1}\le rn$ for $k\le\tau_{rn}^{Z_n}$ and that $N_n<rn$ for n large enough. To check condition (56), note that

$$\begin{aligned}\sup_{t\le T\wedge\tau_r^{Y_n}} \big|B_n(t)-bt\big| &\le \frac{|b|}{n}+\sup_{1\le k\le (Tn)\wedge\tau_{rn}^{Z_n}}\frac{1}{n}\sum_{j=1}^{k}\big|E[Z_{n,j}-Z_{n,j-1}|Z_{n,j-1}]-b\big|\\ &\le \frac{|b|}{n}+\frac{1}{n}\sum_{j=1}^{(Tn)\wedge\tau_{rn}^{Z_n}} f(Z_{n,j-1}\vee N_n)\le\frac{|b|}{n}+T\Big\{\sup_{m\in[N_n,rn]} f(m)\Big\}.\end{aligned}$$

Since this final upper bound is deterministic and vanishes as $n\to\infty$, we have shown that condition (56) holds. Finally, to check condition (57) note that

$$\begin{aligned}\left|A_n(t)-D\int_0^t Y_n^+(s)\,ds\right| &= \left|\frac{1}{n^2}\sum_{k=1}^{\lfloor tn\rfloor}\operatorname{Var}(Z_{n,k}|Z_{n,k-1})-\frac{D}{n^2}\sum_{k=1}^{\lfloor tn\rfloor}Z_{n,k-1}^+-\frac{D}{n}\left(t-\frac{\lfloor tn\rfloor}{n}\right)Z_{n,\lfloor tn\rfloor}^+\right|\\ &\le \frac{1}{n^2}\sum_{\substack{1\le k\le\lfloor tn\rfloor:\\ Z_{n,k-1}>N_n}}\big|\operatorname{Var}(Z_{n,k}|Z_{n,k-1})-DZ_{n,k-1}\big|\\ &\quad+\frac{1}{n^2}\sum_{\substack{1\le k\le\lfloor tn\rfloor:\\ Z_{n,k-1}\le N_n}}\big\{\operatorname{Var}(Z_{n,k}|Z_{n,k-1})+DZ_{n,k-1}^+\big\}+\frac{D}{n}\left(t-\frac{\lfloor tn\rfloor}{n}\right)Z_{n,\lfloor tn\rfloor}^+\\ &\le \frac{1}{n^2}\sum_{\substack{1\le k\le\lfloor tn\rfloor:\\ Z_{n,k-1}>N_n}} g(N_n)Z_{n,k-1}+\frac{tN_n(2D+g(N_n))}{n}+\frac{D}{n}\left(t-\frac{\lfloor tn\rfloor}{n}\right)Z_{n,\lfloor tn\rfloor}^+,\end{aligned} \tag{58}$$

where the last inequality follows from condition (i)(V) and the assumption that g is non-increasing. Now if $t \le \tau_r^{Y_n}$ (equivalently, $tn \le \tau_{rn}^{Z_n}$) then $Z_{n,k-1} < rn$ for all $k \le \lfloor tn \rfloor$, and so the first sum in (58) is at most $rtn^2 g(N_n)$. As for the last term in (58), if $t < \tau_r^{Y_n}$ then $Z_{n,\lfloor tn \rfloor} < rn$ whereas if $t = \tau_r^{Y_n}$ then tn is an integer and the last term in (58) is zero. Thus, we conclude that

$$\sup_{t \le T \wedge \tau_r^{Y_n}} \left| A_n(t) - D \int_0^t Y_n^+(s)\, ds \right| \le rTg(N_n) + \frac{TN_n(2D + g(N_n))}{n} + \frac{Dr}{n} \underset{n\to\infty}{\longrightarrow} 0.$$

This completes the proof of condition (57) and thus also the proof of the lemma. □

Proof of Lemma 6.1. The diffusion approximation is obtained from Lemma 7.1 in two steps: (1) construct a modified process $\bar{V}$ for which it is easy to check the conditions of Lemma 7.1 and conclude the convergence to the diffusion for all times; (2) couple the original process to the modified one so that they coincide up to the first time the processes enter $(-\infty, N_n]$. Since $N_n = o(n)$, this gives the diffusion approximation up to the first entrance time to $(-\infty, n\varepsilon]$ for every $\varepsilon > 0$ as claimed.

Note that the backward BLP V_k can be written as $V_k = \sum_{m=1}^{V_{k-1}+1} \tilde{G}_m^k$, where $\tilde{G}_m^x = F_m^x - F_{m-1}^x$ is the number of failures between the $(m-1)$st and mth success in the sequence of Bernoulli trials $\{\xi_x(j)\}_{j\ge 1}$. It follows that

$$V_k = \sum_{m=1}^{V_{k-1}+1} \tilde{G}_m^k = V_{k-1} + 1 + \sum_{m=1}^{V_{k-1}+1} \big(\tilde{G}_m^k - 1\big), \quad k \ge 1. \tag{59}$$

We now construct the family of modified processes $\bar{V}_n = \{\bar{V}_{n,k}\}_{k\ge 0}$ as follows. Fix any sequence $N_n \in \mathbb{N}$ such that $2 \le N_n \to \infty$ and $N_n = o(n)$ as $n \to \infty$. For all sufficiently large n (we want $N_n \ll ny_n$) set

$$\bar{V}_{n,0} = \lfloor ny_n \rfloor, \quad \text{and} \quad \bar{V}_{n,k} := \bar{V}_{n,k-1} + 1 + \sum_{j=1}^{(\bar{V}_{n,k-1}+1)\vee N_n} \big(\tilde{G}_j^k - 1\big), \quad k \ge 1.$$

Note that the modified process $\bar{V}_n$ is naturally coupled with the BLP V since we use the same sequence of Bernoulli trials in the construction of both. In particular, if we start with $\bar{V}_{n,0} = \lfloor ny_n \rfloor > \varepsilon n > N_n$ then the two process are identical up until exiting $[\varepsilon n, \infty)$.

It remains now to check the conditions of Lemma 7.1 for the family of modified processes $\bar{V}_n$.

Parameters and condition (i). By Propositions 4.1 and 4.2, condition (i)(E) holds for $Z_n = \bar{V}_n$ with $b = 1 + \eta \cdot \tilde{\mathbf{r}}$ and $f(x) = c_{12}e^{-c_{13}x}$, and condition (i)(V) holds with $D = \nu$ and $g(x) = c_{14}/x$.

Condition (ii). Fix $T, r > 0$. We need to show that

$$\lim_{n\to\infty} \frac{1}{n^2} E_\eta \left[\max_{1\le k\le (Tn)\wedge \tau_{rn}^{\bar{V}_n}} \left| 1 + \sum_{j=1}^{(\bar{V}_{n,k-1}+1)\vee N_n} \big(\tilde{G}_j^k - 1\big) \right|^2 \right] = 0,$$

where $\tau_{rn}^{\bar{V}_n} = \inf\{k : \bar{V}_{n,k} \ge rn\}$. To see this, note that if n is large enough so that $N_n < rn$ then

$$\begin{aligned}
\frac{1}{n^2} E_\eta \left(\max_{1\le k\le (Tn)\wedge \tau_{rn}^{\bar{V}_n}} \left| \sum_{j=1}^{(\bar{V}_{n,k-1}+1)\vee N_n} \big(\tilde{G}_j^k - 1\big) \right|^2 \right) &\le \frac{1}{n^2} E_\eta \left(\max_{1\le k\le (Tn)} \max_{N_n\le m\le rn+1} \left| \sum_{j=1}^{m} \big(\tilde{G}_j^k - 1\big) \right|^2 \right) \\
&\le \frac{1}{n^2} \sum_{y=0}^{\infty} P_\eta \left(\max_{1\le k\le (Tn)} \max_{N_n\le m\le rn+1} \left| \sum_{j=1}^{m} \big(\tilde{G}_j^k - 1\big) \right|^2 > y \right) \\
&\le \frac{r^{3/2}}{\sqrt{n}} + rT \sum_{y\ge (rn)^{3/2}} P_\eta \left(\left| \sum_{j=1}^{m} \big(\tilde{G}_j^k - 1\big) \right| > \sqrt{y} \right).
\end{aligned}$$

Finally we apply Lemma A.1 to get that the expression in the last line does not exceed

$$\frac{r^{3/2}}{\sqrt{n}} + rT \sum_{y \geq (rn)^{3/2}} C\left(\exp\left\{-C'\left(\frac{y}{\sqrt{y} \vee (8rn)}\right)\right\} + \exp\{-C'\sqrt{y}\}\right)$$
$$\leq \frac{r^{3/2}}{\sqrt{n}} + rT \sum_{y \geq (rn)^{3/2}} C\left(\exp\left\{-C'\left(\frac{y}{\sqrt{y} \vee (8y^{2/3})}\right)\right\} + \exp\{-C'\sqrt{y}\}\right) \to 0 \quad \text{as } n \to \infty.$$

By Lemma 7.1 we conclude that the processes $t \mapsto \bar{V}_{n,\lfloor nt \rfloor}/n$ with initial conditions $\bar{V}_{n,0} = \lfloor n y_n \rfloor$ and $y_n \to y$ converge in distribution as $n \to \infty$ to the process Y defined by (50). As noted above, since $\bar{V}_n$ can be coupled with V until time $\tau_{\varepsilon n}^V$, this gives us the desired result. □

Next we give the proof of the overshoot estimates in Lemma 6.3 since they are needed for the proof of Lemma 6.2.

Proof of Lemma 6.3. The proof follows closely the arguments in the proof of Lemma 5.1 in [23]. We begin by noting that

$$\max_{0 \leq z < x} P_\eta^{V,z}\left(V_{\tau_x^V} > x + y \mid \tau_x^V < \sigma_0^V\right) \leq \max_{0 \leq z < x} \frac{P_\eta^{V,z}(V_1 > x + y)}{P_\eta^{V,z}(V_1 \geq x)}.$$

Since the (averaged) distribution of the sequence $\{\tilde{G}_j^x\}_{j \geq 1}$ defined above does not depend on $x \in \mathbb{Z}$ we will let $\{\tilde{G}_j\}_{j \geq 1}$ denote a sequence with this same distribution. It follows from (59) that

$$\begin{aligned}\max_{0 \leq z < x} P_\eta^{V,z}\left(V_{\tau_x^V} > x + y \mid \tau_x^V < \sigma_0^V\right) &\leq \max_{0 \leq z < x} \frac{P_\eta(z + 1 + \sum_{j=1}^{z+1}(\tilde{G}_j - 1) > x + y)}{P_\eta(z + 1 + \sum_{j=1}^{z+1}(\tilde{G}_j - 1) \geq x)} \\ &= \max_{0 \leq m < x} \frac{P_\eta(\sum_{j=1}^{x-m}(\tilde{G}_j - 1) > y + m)}{P_\eta(\sum_{j=1}^{x-m}(\tilde{G}_j - 1) \geq m)},\end{aligned} \tag{60}$$

where the last equality follows from the substitution $m = x - z - 1$. We therefore need to give a lower bound on the denominator and an upper bound on the numerator of (60). For the lower bound on the denominator we will use the following lemma.

Lemma 7.2.

$$\inf_{n \geq 1} \min_{i \in \mathcal{R}} P_i\left(\sum_{j=1}^{n}(\tilde{G}_j - 1) \geq 0\right) > 0.$$

Proof. The event $\{\sum_{j=1}^{n}(\tilde{G}_j - 1) \geq 0\}$ occurs when there are at least n failures by the time of the nth success in the sequence of Bernoulli trials. This occurs if and only if there are at least n failures in the first $2n - 1$ trials, or equivalently less than n successes in the first $2n - 1$ trials. That is,

$$P_i\left(\sum_{j=1}^{n}(\tilde{G}_j - 1) \geq 0\right) = P_i\left(\sum_{j=1}^{2n-1}\xi(j) < n\right) = P_i\left(\sum_{j=1}^{2n-1}\left(\xi(j) - \frac{1}{2}\right) \leq 0\right).$$

The last probability converges to $1/2$ for each $i \in \mathcal{R}$ by Lemma A.4. □

Next, we show how to use Lemma 7.2 to obtain a lower bound on the denominator on the right side of (60). For any $m \geq 1$ let $\rho_m = \min\{n \geq 1 : \sum_{j=1}^n (\tilde{G}_j - 1) \geq m\}$. Then, clearly

$$\begin{aligned} P_\eta\left(\sum_{j=1}^{x-m}(\tilde{G}_j - 1) \geq m\right) &= \sum_{n=1}^{x-m} P_\eta\left(\sum_{j=1}^{x-m}(\tilde{G}_j - 1) \geq m, \rho_m = n\right) \\ &\geq \sum_{n=1}^{x-m} P_\eta\left(\sum_{j=n+1}^{x-m}(\tilde{G}_j - 1) \geq 0, \rho_m = n\right) \geq C P_\eta(\rho_m \leq x - m), \end{aligned} \tag{61}$$

where the constant $C > 0$ in the last line comes from Lemma 7.2.

For the numerator in (60),

$$\begin{aligned} &P_\eta\left(\sum_{j=1}^{x-m}(\tilde{G}_j - 1) > y + m\right) \\ &\leq \sum_{n=1}^{x-m} P_\eta\left(\sum_{j=1}^{n}(\tilde{G}_j - 1) > m + \frac{y}{2}, \rho_m = n\right) + \sum_{n=1}^{x-m} P_\eta\left(\sum_{j=n+1}^{x-m}(\tilde{G}_j - 1) > \frac{y}{2}, \rho_m = n\right) \\ &\leq \sum_{n=1}^{x-m}\left\{\max_{i\in\mathcal{R}}\max_{\ell<m} P_i\left(\tilde{G}_1 - 1 > m - \ell + \frac{y}{2} | \tilde{G}_1 - 1 \geq m - \ell\right)\right\} P_\eta(\rho_m = n) \\ &\quad + \sum_{n=1}^{x-m} P_\eta(\rho_m = n)\left\{\max_{i\in\mathcal{R}}\max_{k<x} P_i\left(\sum_{j=1}^{k}(\tilde{G}_j - 1) > \frac{y}{2}\right)\right\} \\ &\leq \left\{p_{\max}^{\lfloor y/2\rfloor} + \max_{i\in\mathcal{R}}\max_{k<x} P_i\left(\sum_{j=1}^{k}(\tilde{G}_j - 1) > \frac{y}{2}\right)\right\} P_\eta(\rho_m \leq x - m), \end{aligned} \tag{62}$$

where $p_{\max} = \max_{i\in\mathcal{R}} p(i) < 1$. Applying (61) and (62) to (60), and using the concentration bounds in Lemma A.1 finishes the proof of the first part of Lemma 6.3. The proof of the second inequality in Lemma 6.3 is essentially the same and is therefore omitted. □

Proof of Lemma 6.2. As noted prior to the statement of Lemma 6.2, if $Y(t)$ is the process in (50) that arises as the scaling limit of the backward BLP, then $Y^\delta(t)$ (or $\log Y(t)$ when $\delta = 0$) is a martingale. The proof of Lemma 6.2 is then accomplished by showing that V_k^δ (or $\log V_k$ when $\delta = 0$) is nearly a martingale prior to exiting the interval (a^{j-1}, a_j). The proof follows essentially the same approach as in [23, proof of Lemma 5.2(ii), pp. 598–599]. To this end, let $s \in C^\infty([0,\infty))$ have compact support, and satisfy

$$s(x) = \begin{cases} x^\delta, & \delta \neq 0, \\ \log x, & \delta = 0 \end{cases} \quad \text{for all } x \in \left(\frac{2}{3a}, \frac{3a}{2}\right).$$

Define $\mathcal{H}_k := \sigma(V_i : i \leq k)$. Then, a Taylor expansion for $s(x)$ yields that on the event $\{\gamma > k\}$

$$E_\eta^{V,x}\left[s\left(\frac{V_{k+1}}{a^j}\right)\Big|\mathcal{H}_k\right] = s\left(\frac{V_k}{a^j}\right) + \frac{s'(\frac{V_k}{a^j})}{a^j} E_\eta^{V,x}[V_{k+1} - V_k|\mathcal{H}_k] + \frac{s''(\frac{V_k}{a^j})}{2a^{2j}} E_\eta^{V,x}\left[(V_{k+1} - V_k)^2|\mathcal{H}_k\right] + \varepsilon_{j,k},$$

where the term $\varepsilon_{j,k}$ comes from the error in the Taylor expansion and can thus be bounded by

$$\varepsilon_{j,k} \leq \frac{\|s'''\|_\infty}{6a^{3j}} E_\eta^{V,x}\left[(V_{k+1} - V_k)^4|\mathcal{H}_k\right]^{3/4} \leq \frac{\|s'''\|_\infty}{6a^{3j}}\left(AV_k^2\right)^{3/4} \leq Ca^{-3j/2}, \tag{63}$$

where the second-to-last inequality follows from the fourth moment bound in Lemma A.3 and the last inequality follows from the fact that $V_k < a^{(j+1)}$ on the event $\{\gamma > k\}$. Next, it follows from Propositions 4.1 and 4.2 that on the event $\{\gamma > k\}$

$$\left|E_\eta^{V,x}[V_{k+1} - V_k|\mathcal{H}_k] - (1 + \eta \cdot \tilde{\mathbf{r}})\right| \le c_{12}e^{-c_{13}V_k} \le c_{12}e^{-c_{13}a^{j-1}} \tag{64}$$

and

$$\left|E_\eta^{V,x}\left[(V_{k+1} - V_k)^2|\mathcal{H}_k\right] - \tilde{\nu}V_k\right| \le C''. \tag{65}$$

Combining (63), (64), and (65) we see that

$$E_\eta^{V,x}\left[s\left(\frac{V_{k+1}}{a^j}\right)\Big|\mathcal{H}_k\right] = s\left(\frac{V_k}{a^j}\right) + \frac{1}{a^j}\left\{(1 + \eta \cdot \tilde{\mathbf{r}})s'\left(\frac{V_k}{a^j}\right) + \frac{\tilde{\nu}}{2}\frac{V_k}{a^j}s''\left(\frac{V_k}{a^j}\right)\right\} + R_{j,k}, \tag{66}$$

where on the event $\{\gamma > k\}$ the error term $R_{j,k}$ is such that

$$|R_{j,k}| \le Ca^{-3j/2} + Ca^{-j}e^{-C'a^j} + Ca^{-2j} \le C''a^{-3j/2}. \tag{67}$$

Now, it follows from (49) and the fact that $s(x) = x^\delta$ in $[1/a, a]$ that

$$(1 + \eta \cdot \tilde{\mathbf{r}})s'(x) + \frac{\tilde{\nu}}{2}xs''(x) = \frac{\tilde{\nu}}{2}\left\{(1 - \delta)s'(x) + xs''(x)\right\} = 0, \quad \forall x \in [1/a, a].$$

Therefore, the quantity inside the braces in (66) vanishes on the event $\{\gamma > k\}$ and so we can conclude that $s(\frac{V_{n\wedge\gamma}}{a^j}) - \sum_{k=0}^{(n\wedge\gamma)-1} R_{j,n}$ is a martingale with respect to the filtration $\mathcal{H}_n$. From this point, the remainder of the proof is the same as that of Lemma 5.2 in [23]. We will give a sketch and refer the reader to [23] for details. First, the diffusion approximation in Lemma 6.1 can be used to show that $E_\eta^{V,x}[\gamma] \le Ca^j$ and thus it follows from the optional stopping theorem and (67) that

$$\left|E_\eta^{V,x}\left[s\left(\frac{V_\gamma}{a^j}\right)\right] - s\left(x/a^j\right)\right| \le E_\eta^{V,x}\left[\left|\sum_{k=0}^{\gamma-1} R_{j,k}\right|\right] \le Ca^{-j/2}. \tag{68}$$

Next, the over/under-shoot estimates in Lemma 6.3 are sufficient to show that

$$\left|E_\eta^{V,x}\left[s\left(\frac{V_\gamma}{a^j}\right)\right] - P_\eta^{V,x}\left(V_\gamma \le a^{j-1}\right)a^{-\delta} - P_\eta^{V,x}\left(V_\gamma \ge a^{j+1}\right)a^\delta\right| \le Ca^{-j/3}. \tag{69}$$

Finally, since $|x - a^j| \le a^{2j/3}$, it follows that $|s(\frac{x}{a^j}) - 1| \le Ca^{-j/3}$. From this and the estimates in (68)–(69) the conclusion of Lemma 6.2 follows. □

Appendix A: Appendix

In this appendix we prove several technical estimates which are used in the analysis of the BLP in Section 7. We begin with the following concentration bound.

Lemma A.1. *Let $\bar{p} = 1/2$. There exist constants $C, C' > 0$ and $y_0 < \infty$ such that for all $n \ge 1$, $y \ge y_0$, and for any initial distribution η for the environment Markov chain we have*

$$P_\eta\left(\left|\sum_{j=1}^n (\tilde{G}_j - 1)\right| > y\right) \le C\left(\exp\left\{-C'\left(\frac{y^2}{y \vee 8n}\right)\right\} + \exp\left\{-C'y\right\}\right).$$

Before giving the proof of Lemma A.1, we note that a similar concentration bound is true for sums of i.i.d. random variables with finite exponential moments.

Lemma A.2 (Theorem III.15 in [34]). *Let $Y_1, Y_2, \dots$ be a sequence of i.i.d. non-negative random variables with $E[Y_1] = \mu$ and $E[e^{\lambda_0 Y_1}] < \infty$ for some $\lambda_0 > 0$. Then, there exists a constant $C > 0$ such that*

$$P\left(\left|\sum_{k=1}^{n} Y_k - \mu n\right| \geq y\right) \leq \exp\left\{-C\frac{y^2}{y \vee n}\right\}.$$

Proof of Lemma A.1. Let i^* be any positive recurrent state, so that $\mu(i^*) > 0$. Fix such an i^* and set $J_0 = \inf\{j \geq 1 : R_j = i^*\}$, and for $k \geq 1$ let $J_k = \inf\{j > J_{k-1} : R_j = i^*\}$. Also, for $k \geq 1$ let

$$Y_k = \sum_{j=J_{k-1}}^{J_k - 1} \xi_j.$$

Note that $\{Y_k\}_{k\geq 1}$ is an i.i.d. sequence with $E_\eta[Y_1] = E_\eta[J_1 - J_0]/2 = 1/(2\mu(i^*))$.

Now, for any $y > 0$ and any integers $n, N \geq 1$

$$P_\eta\left(\sum_{j=1}^{n}(\tilde{G}_j - 1) > y\right) = P_\eta\left(\sum_{j=1}^{2n+y} \xi_j < n\right)$$
$$\leq P_\eta\left(J_0 > \frac{y}{2}\right) + P_\eta\left(J_N - J_0 > 2n + \frac{y}{2}\right) + P_\eta\left(\sum_{k=1}^{N} Y_k < n\right).$$

Now, let $N = \lfloor \mu(i^*)(\frac{y}{4} + 2n) \rfloor$ so that

$$2n + \frac{y}{2} \geq N E[J_1 - J_0] + \frac{y}{4} \quad \text{and} \quad n \leq N E[Y_1] - \frac{y}{8} + \frac{1}{2\mu(i^*)}.$$

Therefore, for $y > 0$ sufficiently large (so that $y/8 - 1/(2\mu(i^*)) > y/9$; thus $y > 36/\mu(i^*)$ is sufficient) we have

$$P_\eta\left(\sum_{j=1}^{n}(\tilde{G}_j - 1) > y\right) \leq P_\eta\left(J_0 > \frac{y}{2}\right) + P_\eta\left(J_N - J_0 > N E[J_1 - J_0] + \frac{y}{4}\right) + P_\eta\left(\sum_{k=1}^{N} Y_k < N E[Y_1] - \frac{y}{9}\right).$$

Since the cookie stack $\{\omega_0(j)\}_{j\geq 1}$ is a finite state Markov chain and i^* is in the unique irreducible subset, then it follows easily that there exists constants $C_1, C_2 > 0$ such that

$$P_\eta\left(J_0 > \frac{y}{2}\right) \leq C_1 e^{-C_2 y},$$

for all $y > 0$ and for all distributions η of the environment Markov chain. From this it also follows that $J_{k+1} - J_k$ has exponential tails, and since $Y_k \leq J_{k+1} - J_k$ then Y_k also has exponential tails. Therefore, we can apply Lemma A.2 to get that

$$P_\eta\left(J_N - J_0 > N E[J_1 - J_0] + \frac{y}{4}\right) + P_\eta\left(\sum_{k=1}^{N} Y_k < N E[Y_1] - \frac{y}{9}\right)$$
$$\leq e^{-C_3 \frac{y^2/16}{N \vee (y/4)}} + e^{-C_4 \frac{y^2/81}{N \vee (y/9)}} \leq 2 e^{-C_5 \frac{y^2}{(9N) \vee y}}.$$

Now, recalling the definition of N (which depends on both n and y) we see that $N \le \mu(i^*)(8n \vee y)/2 < (8n \vee y)/2$. Therefore, we can conclude that

$$P_\eta\left(\sum_{j=1}^n (\tilde{G}_j - 1) > y\right) \le C_1 e^{-C_2 y} + 2e^{-C_6 \frac{y^2}{(8n)\vee y}},$$

for all $n \ge 1$ all $y \ge \lceil 36/\mu(i^*)\rceil =: y_0$.

The proof of the upper bound for the left tails is similar. Indeed,

$$P_\eta\left(\sum_{j=1}^n (\tilde{G}_j - 1) < -y\right) = P_\eta\left(\sum_{j=1}^n \tilde{G}_j < n - y\right).$$

If $y \ge n$ then the probability is 0 and there is nothing to prove. Assume that $n > y$. Then the last probability is equal to

$$P_\eta\left(\sum_{j=1}^{2n-y-1} (1 - \xi_i) < n - y\right) = P_\eta\left(\sum_{j=1}^{2(n-y)+y-1} (1 - \xi_i) < (n - y)\right).$$

This probability can be handled in exactly the same way as the upper bound for the right tails by setting $n' = n - y$. Then in the exponent we shall have (for $y < n$) $-y^2/(8(n-y) \vee y) \le -y^2/(8n) = -y^2/(8n \vee y)$. $\square$

The concentration bounds in Lemma A.1 can be used to prove the following moment bound for sums of the $\tilde{G}_i$.

Lemma A.3. *Let $\bar{p} = 1/2$. Then there is a constant $A = A(y_0, C, C') > 0$ (see Lemma A.1) such that for any initial distribution η on the environment Markov chain and all $n \ge 1$*

$$E_\eta\left[\left(\sum_{i=1}^n (\tilde{G}_i - 1)\right)^4\right] \le An^2.$$

Proof. We shall use the fact that for a non-negative integer-valued random variable X

$$E(X^4) \le 1 + c\sum_{y=1}^\infty y^3 P(X > y).$$

Set $X = |\sum_{i=1}^n (\tilde{G}_i - 1)|$. Then by Lemma A.1 for all $n > y_0/8$

$$\begin{aligned}
E(X^4) &\le 1 + c\sum_{y=1}^{y_0-1} y^3 P(X > y) + C'' \sum_{y=y_0}^\infty y^3 \left(e^{-C'y^2/(8n\vee y)} + e^{-C'y}\right) \\
&\le K_1(y_0, C, C') + C'' \sum_{y=y_0}^{8n} y^3 e^{-C'y^2/(8n)} + 2C'' \sum_{y=y_0}^\infty y^3 e^{-C'y} \\
&\le K_2(y_0, C, C') + C'' \sum_{y=y_0}^{8n} y^3 e^{-C'y^2/(8n)}.
\end{aligned}$$

Since the function $y^3 e^{-C'y^2/8}$ is directly Riemann integrable, we can bound the sum in the last line by n^2 times a Riemann sum approximation. That is,

$$\sum_{y=y_0}^{8n} y^3 e^{-C'y^2/(8n)} \le n^2 \sum_{y=0}^\infty \left(\frac{y}{\sqrt{n}}\right)^3 e^{-\frac{C'}{8}(\frac{y}{\sqrt{n}})^2} \frac{1}{\sqrt{n}} \sim n^2 \int_0^\infty y^3 e^{-C'y^2/8}\,dy, \quad \text{as } n \to \infty.$$

$\square$

Lemma A.4. *Assume that $\bar{p} = 1/2$. There exists a constant $v > 0$ such that for any initial distribution η, under the measure P_η we have that $\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i - 1/2) \Rightarrow vZ$, where Z is a standard normal random variable.*

Proof. As in the proof of Lemma A.1, fix a recurrent state $i^* \in \mathcal{R}$ and let $J_0, J_1, J_2, \dots$ be the successive return times of the Markov chain to i^*. Set

$$Z_k := \sum_{j=J_{k-1}}^{J_k - 1} (\xi_j - 1/2), \quad k \geq 1.$$

$(Z_k)_{k\geq 1}$ is an i.i.d. sequence of centered square integrable random variables. Let $L_n = 0$ if $J_0 > n$ and $L_n = \max\{k \in \mathbb{N} : J_{k-1} \leq n\}$ otherwise. Then

$$\begin{aligned}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^n(\xi_i - 1/2) - \sum_{k=1}^{L_n} Z_k\right| &= \frac{1}{\sqrt{n}}\left|\sum_{i=1}^{J_0-1}(\xi_i - 1/2) - \sum_{i=n+1}^{L_{n+1}-1}(\xi_i - 1/2)\right| \\ &\leq \frac{1}{2\sqrt{n}}\Big(J_0 + \max_{1\leq k\leq n+1}(J_k - J_{k-1})\Big).\end{aligned}$$

Since J_0 is almost surely finite and $\{J_k - J_{k-1}\}_{k\geq 1}$ is i.i.d. with finite second moment, it follows that the last expression above converges to 0 in probability. Therefore, it remains only to prove that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{L_n} Z_k \Rightarrow vZ$$

for some $v > 0$. By the law of large numbers, $L_n/n \to 1/E_{i^*}[J_0] = \mu(i^*)$ a.s., and the desired conclusion follows from [15, Theorem I.3.1(ii), p. 17] with $v^2 = \mu(i^*)E_\eta[Z_1^2]$. □

Appendix B: Details of the limiting distribution in the case $\delta = 2$

In this appendix we will give the details of the proof of the limiting distributions in Theorem 1.8 in the case $\delta = 2$. This is the most subtle case in Theorem 1.8 and has not been fully written down for ERWs even for the model with boundedly many cookies per site. Even though the proof is very similar to that for one-dimensional random walk in random environment for the analogous case ([21, pp. 166–168]), some of the details are different due to the fact that the regeneration times r_i have infinite second moment.

First, recall the definition (1) of the α-stable distribution $L_{\alpha,b}$ and the fact that if Z has distribution $L_{\alpha,b}$ with $\alpha \neq 1$, then cZ has distribution L_{α,bc^α} for any $c > 0$. However, if $\alpha = 1$ then a re-centering is needed to get a distribution of the same form. This difference is indicative of the fact that for $\alpha \neq 1$ there is a natural choice of the centering for totally asymmetric α-stable distributions: the distributions $L_{\alpha,b}$ have mean zero when $\alpha > 1$ and have support equal to $[0, \infty)$ when $\alpha < 1$. In contrast, when $\alpha = 1$ there is no canonical choice of the centering for the totally asymmetric 1-stable distributions. To this end, for $b > 0$ and $\xi \in \mathbb{R}$ let $L_{1,b,\xi}$ be the probability distribution with characteristic exponent given by

$$\log \int_{\mathbb{R}} e^{iux} L_{1,b,\xi}(dx) = iu\xi - b|u|\left(1 + \frac{2i}{\pi}\log|u|\operatorname{sign}(u)\right).$$

That is, the distribution $L_{1,b,\xi}$ differs from $L_{1,b}$ by a simple spatial translation.

Recall that r_k is the time of the kth return of the backward BLP V_i to zero and that $W_k = \sum_{i=r_{k-1}}^{r_k - 1} V_i$. When $\delta = 2$, it follows from Theorem 2.7 that

$$P_\eta^{V,0}(r_1 > t) \sim c_3 t^{-2} \quad \text{and} \quad P_\eta^{V,0}(W_1 > t) \sim c_4 t^{-1}, \quad \text{as } t \to \infty, \tag{70}$$

for some constants $c_3 = c_3(0,\eta) > 0$ and $c_4 = c_4(0,\eta) > 0$. Let $m(t) = E_\eta^{V,0}[W_1 \mathbf{1}_{\{W_1 \le t\}}]$ be the truncated first moment of W_1 (note that the above tail asymptotics of W_1 imply that $m(t) \sim c_4 \log t$). Then, it follows from [10, Theorem 3.7.2] or [14, Theorem 17.5.3] that there exist constants $b' > 0$ and $\xi' \in \mathbb{R}$ such that

$$\lim_{n\to\infty} P_\eta^{V,0}\left(\frac{\sum_{k=1}^n W_k - nm(n)}{n} \le x \right) = L_{1,b',\xi'}(x), \quad x \in \mathbb{R}, \tag{71}$$

and that there exists a constant $A > 0$ such that

$$\lim_{n\to\infty} P_\eta^{V,0}\left(\frac{r_n - n\bar{r}}{A\sqrt{n\log n}} \le x \right) = \Phi(x), \quad x \in \mathbb{R}. \tag{72}$$

Since (72) implies that $r_{\lfloor n/\bar{r}\rfloor}$ is typically close to n, we wish to approximate the distribution of $n^{-1}\sum_{i=1}^n V_i$ by that of $n^{-1}\sum_{k=1}^{n/\bar{r}} W_k$. Indeed, we claim that the difference converges to zero in $P_\eta^{V,0}$-probability. To see this, note that for any $\varepsilon > 0$,

$$\begin{aligned} P_\eta^{V,0}\left(\left| \sum_{i=1}^n V_i - \sum_{k=1}^{n/\bar{r}} W_k \right| \ge \varepsilon n \right) &\le P_\eta^{V,0}\big(|r_{\lfloor n/\bar{r}\rfloor} - n| > n^{3/4}\big) + P_\eta^{V,0}\left(\sum_{|k-\frac{n}{\bar{r}}| \le n^{3/4}+1} W_k \ge \varepsilon n \right) \\ &= P_\eta^{V,0}\big(|r_{\lfloor n/\bar{r}\rfloor} - n| > n^{3/4}\big) + P_\eta^{V,0}\left(\sum_{k=1}^{2\lfloor n^{3/4}\rfloor+3} W_k \ge \varepsilon n \right). \end{aligned}$$

It follows from (71) and (72) that both terms in the last line above vanish as $n \to \infty$ for any $\varepsilon > 0$. Therefore, we can conclude from this, the limiting distribution in (71), and the fact that T_n has the same limiting distribution as $n + 2\sum_{i=1}^n V_i$ that

$$\begin{aligned} \lim_{n\to\infty} P_\eta^{V,0}\left(\frac{T_n - 2\frac{n}{\bar{r}} m(\frac{n}{\bar{r}})}{n} \le x \right) &= \lim_{n\to\infty} P_\eta^{V,0}\left(\frac{n + 2\sum_{i=1}^n V_i - 2\frac{n}{\bar{r}} m(\frac{n}{\bar{r}})}{n} \le x \right) \\ &= \lim_{n\to\infty} P_\eta^{V,0}\left(\frac{\sum_{k=1}^{\lfloor n/\bar{r}\rfloor} W_k - \frac{n}{\bar{r}} m(\frac{n}{\bar{r}})}{n/\bar{r}} \le \frac{(x-1)\bar{r}}{2} \right) \\ &= L_{1,b',\xi'}\left(\frac{(x-1)\bar{r}}{2} \right) = L_{1,b,\xi}(x), \end{aligned}$$

where in the last equality we have $b = \frac{2b'}{\bar{r}}$ and $\xi = 1 + \frac{2\xi'}{\bar{r}} - \frac{4b'}{\pi\bar{r}} \log(\frac{2}{\bar{r}})$. This completes the proof of the limiting distribution for T_n when $\delta = 2$ with $b > 0$ as above, $D(n) = \xi + \frac{2}{\bar{r}} m(\frac{n}{\bar{r}})$, and $a = \bar{r}/(2c_4)$.

Before proving the limiting distribution for X_n, we first need to remark on the specific choice of the centering term $D(n)$ in the limiting distribution for T_n. One cannot use an arbitrary centering term growing asymptotically like $a^{-1}\log n$, but the above choice of the centering term by the function $D(t) = \xi + \frac{2}{\bar{r}} E[W_1 \mathbf{1}_{\{W_1 \le t/\bar{r}\}}]$ grows regularly enough due to the tail asymptotics of W_1 so that

$$\lim_{n\to\infty} D(m_n) - D(n) = 0, \quad \text{if } m_n \sim n \text{ as } n \to \infty. \tag{73}$$

For $t > 0$ let $\Gamma(t) = \inf\{s > 0 : sD(s) \ge t\}$. Then, it follows that

$$\Gamma(t) \sim \frac{at}{\log(t)} \quad \text{and} \quad \Gamma(t)D\big(\Gamma(t)\big) = t + o\big(\Gamma(t)\big), \quad \text{as } t \to \infty. \tag{74}$$

The first asymptotic expression in (74) follows easily from the definition of $\Gamma(t)$ and the fact that $D(t) \sim a^{-1}\log t$. For the second asymptotic expression in (74), note that for s sufficiently large $sD(s)$ is strictly increasing and right continuous in s. Moreover, if $sD(s)$ is discontinuous at s_0 then the size of the jump discontinuity is $2(s_0/\bar{r})^2 P(W_1 = s_0/\bar{r})$. Therefore,

$$\bigl|\Gamma(t)D\bigl(\Gamma(t)\bigr) - t\bigr| \le 2\left(\frac{\Gamma(t)}{\bar{r}}\right)^2 P\left(W_1 = \frac{\Gamma(t)}{\bar{r}}\right),$$

since if $sD(s)$ is continuous at $s = \Gamma(t)$ then the above difference is zero while if $sD(s)$ is discontinuous at $s = \Gamma(t)$ then the difference is at most the size of the jump discontinuity at that point. Since the tail asymptotics of W_1 imply that $xP(W_1 = x) = o(1)$, the second asymptotic expression in (74) follows.

Now, for any $x \in \mathbb{R}$ and $n \ge 1$ let $m_{n,x} := \lceil \Gamma(n) + \frac{xn}{(\log n)^2} \rceil \vee 0$. Since $m_{n,x}$ grows asymptotically like $\Gamma(n)$ as $n \to \infty$ it follows that

$$\begin{aligned}\lim_{n\to\infty} \frac{n - m_{n,x} D(m_{n,x})}{m_{n,x}} &= \lim_{n\to\infty} \frac{n - m_{n,x} D(\Gamma(n))}{m_{n,x}} \\ &= \lim_{n\to\infty} \frac{n - (\Gamma(n) + \frac{xn}{(\log n)^2}) D(\Gamma(n))}{\Gamma(n) + \frac{xn}{(\log n)^2}} = \lim_{n\to\infty} \frac{-\frac{xn}{(\log n)^2} D(\Gamma(n))}{\Gamma(n) + \frac{xn}{(\log n)^2}} = \frac{-x}{a^2},\end{aligned} \tag{75}$$

where the first equality follows from (73) and the last two equalities follow from (74). Similarly, letting $M_{n,x} = m_{n,x} + \lceil \sqrt{n} \rceil$ it follows that

$$\lim_{n\to\infty} \frac{n - M_{n,x} D(M_{n,x})}{M_{n,x}} = \frac{-x}{a^2}. \tag{76}$$

Finally, it follows from (13) and (14) that

$$\begin{aligned}&P_\eta\left(\frac{T_{m_{n,x}} - m_{n,x} D(m_{n,x})}{m_{n,x}} > \frac{n - m_{n,x} D(m_{n,x})}{m_{n,x}}\right) \\ &\quad = P_\eta(T_{m_{n,x}} > n) \le P_\eta\left(\frac{X_n - \Gamma(n)}{n/(\log n)^2} < x\right) \le P_\eta(T_{M_{n,x}} > n) + \mathcal{O}\bigl(n^{-1/2}\bigr) \\ &\quad = P_\eta\left(\frac{T_{M_{n,x}} - M_{n,x} D(M_{n,x})}{M_{n,x}} > \frac{n - M_{n,x} D(M_{n,x})}{M_{n,x}}\right) + \mathcal{O}\bigl(n^{-1/2}\bigr).\end{aligned}$$

From the limiting distribution for T_n, together with (75) and (76), we conclude that the first and the last probabilities in the display above converge to $1 - L_{1,b}(-x/a^2)$.

Acknowledgements

Elena Kosygina thanks Institut Mittag-Leffler for support and stimulating research environment.

References

[1] G. Amir, N. Berger and T. Orenshtein. Zero–one law for directional transience of one dimensional excited random walks. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (1) (2016) 47–57. MR3449293

[2] G. Amir and T. Orenshtein. Excited Mob. *Stochastic Process. Appl.* **126** (2) (2016) 439–469. MR3434990

[3] A.-L. Basdevant and A. Singh. On the speed of a cookie random walk. *Probab. Theory Related Fields* **141** (3–4) (2008) 625–645. MR2391167

[4] A.-L. Basdevant and A. Singh. Rate of growth of a transient cookie random walk. *Electron. J. Probab.* **13** (26) (2008) 811–851. MR2399297

[5] I. Benjamini and D. B. Wilson. Excited random walk. *Electron. Commun. Probab.* **8** (2003) 86–92. (electronic). MR1987097

[6] K. L. Chung. *Markov Chains with Stationary Transition Probabilities*, 2nd edition. *Die Grundlehren der Mathematischen Wissenschaften* **104**. Springer, New York, 1967. MR0217872

[7] D. Dolgopyat. Central limit theorem for excited random walk in the recurrent regime. *ALEA Lat. Am. J. Probab. Math. Stat.* **8** (2011) 259–268. MR2831235

[8] D. Dolgopyat and E. Kosygina. Scaling limits of recurrent excited random walks on integers. *Electron. Commun. Probab.* **17** (35) (2012) 14. MR2965748

[9] D. Dolgopyat and E. Kosygina. Excursions and occupation times of critical excited random walks. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** (1) (2015) 427–450. MR3368965

[10] R. Durrett. *Probability: Theory and Examples*, 4th edition. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. MR2722836

[11] N. Eisenbaum, H. Kaspi, M. B. Marcus, J. Rosen and Z. Shi. A Ray–Knight theorem for symmetric Markov processes. *Ann. Probab.* **28** (4) (2000) 1781–1796. MR1813843

[12] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York, 1986. MR0838085

[13] W. Feller. Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* 227–246. University of California Press, Berkeley and Los Angeles, 1950. 1951. MR0046022

[14] W. Feller. *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd edition. Wiley, New York, 1971. MR0270403

[15] A. Gut. *Stopped Random Walks: Limit Theorems and Applications*, 2nd edition. *Springer Series in Operations Research and Financial Engineering*. Springer, New York, 2009. MR2489436

[16] T. E. Harris. First passage and recurrence distributions. *Trans. Amer. Math. Soc.* **73** (1952) 471–486. MR0052057

[17] M. Holmes. On strict monotonicity of the speed for excited random walks in one dimension. *Electron. Commun. Probab.* **20** (41) (2015) 7. MR3358963

[18] M. Holmes and T. S. Salisbury. A combinatorial result with applications to self-interacting random walks. *J. Combin. Theory Ser. A* **119** (2) (2012) 460–475. MR2860606

[19] P. Jagers. *Branching Processes with Biological Applications*. *Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics*. Wiley, New York, 1975. MR0488341

[20] M. Jiřina. On Feller's branching diffusion processes. *Čas. Pěst. Mat.* **94** (107) (1969) 84–90. MR0247676

[21] H. Kesten, M. V. Kozlov and F. Spitzer. A limit law for random walk in a random environment. *Compos. Math.* **30** (1975) 145–168. MR0380998

[22] F. B. Knight. Random walks and a sojourn density process of Brownian motion. *Trans. Amer. Math. Soc.* **109** (1963) 56–86. MR0154337

[23] E. Kosygina and T. Mountford. Limit laws of transient excited random walks on integers. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** (2) (2011) 575–600. MR2814424

[24] E. Kosygina and M. Zerner. Excited random walks: Results, methods, open problems. *Bull. Inst. Math. Acad. Sin. (N.S.)* **8** (1) (2013) 105–157. MR3097419

[25] E. Kosygina and M. P. W. Zerner. Positively and negatively excited random walks on integers, with branching processes. *Electron. J. Probab.* **13** (64) (2008) 1952–1979. MR2453552

[26] E. Kosygina and M. P. W. Zerner. Excursions of excited random walks on integers. *Electron. J. Probab.* **19** (25) (2014) 1–25. MR3174837

[27] G. Kozma, T. Orenshtein and I. Shinkar. Excited random walk with periodic cookies. *Ann. Inst. Henri Poincaré Probab. Stat.* (2014). To appear.

[28] T. G. Kurtz. *Approximation of Population Processes*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **36**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1981. MR0610982

[29] M. B. Marcus and J. Rosen. *Markov Processes, Gaussian Processes, and Local Times*. *Cambridge Studies in Advanced Mathematics* **100**. Cambridge University Press, Cambridge, 2006. MR2250510

[30] T. Mountford, L. P. R. Pimentel and G. Valle. On the speed of the one-dimensional excited random walk in the transient regime. *ALEA Lat. Am. J. Probab. Math. Stat.* **2** (2006) 279–296. MR2285733

[31] J. Peterson. Large deviations and slowdown asymptotics for one-dimensional excited random walks. *Electron. J. Probab.* **17** (48) (2012) 24. MR2946155

[32] J. Peterson. Strict monotonicity properties in one-dimensional excited random walks. *Markov Process. Related Fields* **19** (4) (2013) 721–734. MR3185198

[33] J. Peterson. Extreme slowdowns for one-dimensional excited random walks. *Stochastic Process. Appl.* **125** (2) (2015) 458–481. MR3293290

[34] V. V. Petrov. *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* **82**. Springer, New York, 1975. MR0388499

[35] R. G. Pinsky. Transience/recurrence and the speed of a one-dimensional random walk in a "have your cookie and eat it" environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (4) (2010) 949–964. MR2744879

[36] R. G. Pinsky and N. F. Travers. Transience, recurrence and the speed of a random walk in a site-based feedback environment. *Probab. Theory Related Fields* (2016) 1–62.

[37] A.-S. Sznitman. Slowdown estimates and central limit theorem for random walks in random environment. *J. Eur. Math. Soc. (JEMS)* **2** (2) (2000) 93–143. MR1763302

[38] A.-S. Sznitman. *Topics in Occupation Times and Gaussian Free Fields*. *Zurich Lectures in Advanced Mathematics*. European Mathematical Society (EMS), Zürich, 2012. MR2932978

[39] B. Tóth. The "true" self-avoiding walk with bond repulsion on $\mathbf{Z}$: Limit theorems. *Ann. Probab.* **23** (4) (1995) 1523–1556. MR1379158

[40] B. Tóth. Generalized Ray-Knight theory and limit theorems for self-interacting random walks on $\mathbf{Z}^1$. *Ann. Probab.* **24** (3) (1996) 1324–1367. MR1411497

[41] B. Tóth. Self-interacting random motions – a survey. In *Random Walks* 349–384. *Budapest*, *1998*. *Bolyai Soc. Math. Stud.* **9**. János Bolyai Math. Soc., Budapest, 1999. MR1752900

[42] M. P. W. Zerner. Multi-excited random walks on integers. *Probab. Theory Related Fields* **133** (1) (2005) 98–122. MR2197139