Probab. Theory Relat. Fields 91, 195–249 (1992)

Probability Theory and Related Fields

Branching random walk in random environment: phase transitions for local and global growth rates

Andreas Greven[1] **and Frank den Hollander**[2]

[1] Institut für Mathematische Stochastik, Universität Göttingen, Lotzestrasse 13, W-3400 Göttingen, Federal Republic of Germany
[2] Rijksuniversiteit te Utrecht, Budapestlaan 6, 3508 TA Utrecht, The Netherlands

Received August 23, 1990; in revised form August 27, 1991

Summary. Let (η_n) be the infinite particle system on $\mathbb{Z}$ whose evolution is as follows. At each unit of time each particle independently is replaced by a new generation. The size of a new generation descending from a particle at site x has distribution F_x and each of its members independently jumps to site $x \pm 1$ with probability $(1 \pm h)/2$, $h \in [0, 1]$. The sequence $\{F_x\}$ is i.i.d. with uniformly bounded second moment and is kept fixed during the evolution. The initial configuration η_0 is shift invariant and ergodic.

Two quantities are considered:

(1) the global particle density D_n
(= large volume limit of number of particles per site at time n);

(2) the local particle density d_n
(= average number of particles at site 0 at time n).

We calculate the limits ρ and λ of $n^{-1} \log(D_n)$ and $n^{-1} \log(d_n)$ explicitly in the form of two variational formulas. Both limits (and variational formulas) do not depend on the realization of $\{F_x\}$ a.s. By analyzing the variational formulas we extract how ρ and λ depend on the drift h for fixed distribution of F_x. It turns out that the system behaves in a way that is drastically different from what happens in a spatially homogeneous medium:

(i) Both $\rho(h)$ and $\lambda(h)$ exhibit a phase transition associated with localization vs. delocalization at two respective critical values h_1 and h_3 in $(0, 1)$. Here the behavior of the path of descent of a typical particle in the whole population resp. in the population at 0 changes from moving on scale $o(n)$ to moving on scale n. We extract variational expressions for h_1 and h_3.

(ii) Both $\rho(h)$ and $\lambda(h)$ change sign at two respective critical values h_2 and h_4 in $(0, 1)$ (for suitable distribution of F_x). That is, the system changes from survival to extinction on a global resp. on a local scale.

(iii) $\rho(h) \geqq \lambda(h)$ for all h; $\rho(h) = \lambda(h)$ for h sufficiently small and $\rho(h) > \lambda(h)$ for h sufficiently large. This means that the system develops a clustering phenomenon as h increases: the population has large peaks on a thin set.

(iv) $\rho(h) > 0 > \lambda(h)$ for a range of h. (extreme clustering of the system)

We formulate certain technical properties of the variational formulas that are needed in order to derive the qualitative picture of the phase diagram in its full glory. The proof of these properties is deferred to a forthcoming paper dealing exclusively with functional analytic aspects.

The variational formulas reveal a selection mechanism: the typical particle has a path of descent that is best adapted to the given $\{F_x\}$ and that is atypical under the law of the underlying random walk. The random medium induces "selection of the fittest".

0 Introduction and main results

0a Introduction

Infinite particle systems evolving in a random environment, i.e. systems of locally interacting components with a *spatially inhomogeneous* evolution mechanism, exhibit interesting new phenomena not present in their spatially homogeneous counterparts. Although this area is still largely unexplored and only few models have so far been investigated, it has become clear that new intuition and new techniques are needed to understand the nature of these effects. For reference see: Dawson and Fleischmann [4], [5], [6]; Greven [12], [13], [14]; Bramson et al. [2]; Ferreira [10]; Gärtner and Molchanov [11]; Chen and Liggett [3]; Liggett [17]; Baillon et al. [0]. A common feature of all of these papers is that the central question reduces, via some duality relation or some comparison techniques, to a problem involving a single particle in a random environment. It is after this reduction that the real work starts. Therefore these papers stand in the tradition of the literature on classical single particle models in random media, although the problems that come up are of a rather different nature.

In the present paper we study a model for population growth in random media. The new aspect of our approach is that we extract *explicit variational formulas*. From these we deduce a very rich *phase diagram* as a function of a single parameter.

We shall be interested in branching random walks on $\mathbb{Z}$ where the branching mechanism is spatially inhomogeneous and the underlying random walk kernel has a drift. That is, particles perform independent random walks and along the way branch according to a law depending on their location. It was observed by Greven [13] that in systems of such type the drift of the underlying random walk kernel plays an important role for the long term behavior of the population density. In fact, it was conjectured that a phase transition occurs as the drift varies: for suitable choice of parameters describing the environment the population grows when the drift is small but dies out when the drift is large.

We shall study both the *local* and the *global population density*. By the latter we mean the large volume limit of the number of particles per site. We shall show that both these densities grow or decay exponentially and we shall calculate their growth rates explicitly by giving a representation in the form of two variational formulas. These formulas will tell us that three significant changes occur from the spatially homogeneous model (see Fig. 1): (1) If we randomly pick a particle at time n (drawn either locally or globally from the popula-

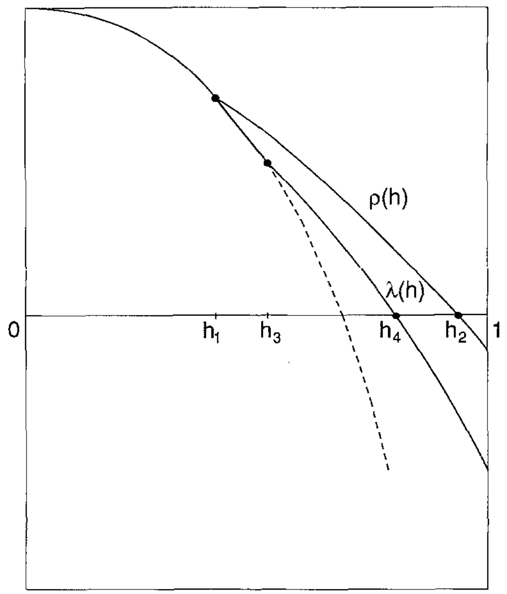

Fig. 1. Qualitative picture of the phase diagram for $\rho(h)$ and $\lambda(h)$, the global resp. local exponential growth rate of the particle density as a function of the drift h. The dashed curve is $\log[M(1-h^2)^{1/2}]$. The end points are $\log \int b\beta(\mathrm{d}b)$ resp. $\int \log b\beta(\mathrm{d}b)$ (see sect. 0d)

tion), then its history will be very different from that in the homogeneous situation. The random environment causes a selection very much following a "survival of the fittest" principle: the population will largely consist of those particles whose ancestors had a pattern of migration that was *best adapted* to the given environment. For example, if with each particle we associate its *path of descent*, then it will turn out that for a typical particle this path moves at a speed different from the drift of the underlying random walk kernel. In fact, for drift below two respective critical values it only moves on scale $o(n)$, a phenomenon which we call *localization*, while beyond the critical value it moves on scale n, which we call *delocalization*; (2) Both the local and the global exponential growth rate may exhibit another phase transition as the drift increases, changing from positive to negative at two respective critical values. That is, the system changes from *survival* to *extinction*. This can only happen when the environment has both super and subcritical offspring distributions. (3) Interesting enough, the two growth rates are *equal* or are *different depending on the drift*. In the latter situation the population experiences *clustering*. In particular, the critical points from point two above may be different and we may have the following remarkable picture: at small drift the population grows everywhere, at intermediate drift it locally dies out but globally grows, and at large drift it dies out both locally and globally. This says that there is an *intermediate phase* where the particles exhibit *extreme clustering* and the population concentrates on a thinning set carrying a growing amount of particles. All these phenomena are absent in the homogeneous model, where the two growth rates not only are identical but also do not depend on the drift.

The study of our system proceeds via a duality relation expressing the two densities of interest in terms of a functional of a single *random walk in random* *scenery*. From this relation, via a sequence of manipulations and combinatorial estimates, we reduce the problem to the study of a quantity of the form

$$\frac{1}{n} \log E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]} f_x(l(x))\right] \chi\left\{\sum_{x=0}^{[\theta n]} l(x)=n\right\}\right).$$

Here f_x is a fixed convex function in the global case and a random linear function depending on the environment at site x in the local case, $\{l(x)\}_{x \geqq 0}$ is the sequence of local times of the random walk, E_θ denotes expectation with respect to the random walk with drift $\theta \in (0, 1)$, and χ denotes indicator. Via the observation that $\{l(x)\}_{x \geqq 0}$ is a two-block functional of a specific Markov sequence (on infinite state space) associated with the random walk, we are then able to characterize the limiting behavior of this quantity in terms of a variational formula using large deviation techniques (not quite in a standard form). By subsequently analyzing the two variational formulas we deduce the above described phase transitions. In addition, the variational formulas allow for a nice interpretation in terms of the branching random walk: they give us information about the path of descent of a typical particle.

Even though our analysis makes use of special properties of f_x and $\{l(x)\}_{x \geqq 0}$, it has various new aspects that make it interesting in the broader context of Markov chains in random scenery. In particular, we stress the role of empirical pair distributions along the level sets of the random scenery.

In the rest of this section we define the model, state the two main theorems and their consequences for the phase diagram, give extensions and finally formulate some conjectures.

0b The model

We now define a system of discrete time branching random walks on $\mathbb{Z}$ evolving in a spatially inhomogeneous environment which is chosen randomly and is kept fixed as the system evolves. To do so we need the following ingredients:

(i) A *random environment* $F = \{F_x\}_{x \in \mathbb{Z}}$, given by a sequence of i.i.d. random variables taking values in the set of probability measures on $\mathbb{N} \cup \{0\}$. F_x plays the role of offspring distribution for particles at site x. We focus on the case

$$\begin{aligned} &P(F_x = F_1) = q \\ &P(F_x = F_2) = 1 - q \end{aligned} \tag{0.1}$$

with F_1 and F_2 two different given offspring distributions satisfying

$$\begin{aligned} &\sum_{n=0}^{\infty} n F_1(n) = b_1 \\ &\sum_{n=0}^{\infty} n F_2(n) = b_2 \\ &0 < \sum_{n=0}^{\infty} n^2 [F_1(n) + F_2(n)] < \infty. \end{aligned} \tag{0.2}$$

This simple form of the environment is not really necessary for our treatment. However, since it makes our results and statements more transparent, we shall stick with it for a while and defer extension to Sect. 0d.

(ii) A *homogeneous transition kernel* $p_h(x, y)$ on $\mathbb{Z} \times \mathbb{Z}$, given by

$$p_h(x, y) = \tfrac{1}{2}(1 \pm h) \quad \text{for } y = x \pm 1, \text{zero otherwise.} \tag{0.3}$$

$p_h(x, y)$ is the transition kernel of the random walk controlling the motion of particles, and h is the *drift parameter.*

(iii) An *initial configuration* $\eta_0 = \{\eta_0(x)\}_{x \in \mathbb{Z}}$, distributed according to a shift invariant and ergodic probability measure on $(\mathbb{N} \cup \{0\})^{\mathbb{Z}}$ with

$$\begin{aligned} &E(\eta_0(x)) = \gamma \\ &E(\eta_0^2(x)) < \infty. \end{aligned} \tag{0.4}$$

η_0 is independent of the environment and $\eta_0(x)$ represents the number of particles at site x at time 0.

For *fixed F* our branching random walk

$$\eta_n = \{\eta_n(x)\}_{x \in \mathbb{Z}} \qquad (n \in \mathbb{N} \cup \{0\})$$

is now defined as the Markov chain on $n(\mathbb{N} \cup \{0\})^{\mathbb{Z}}$ starting in η_0, where at each unit of time each particle is independently replaced by a new generation. The size of a new generation descending from a particle at site x is distributed according to F_x, and the members are located independently according to $p_h(x, y)$. Formally,

$$\eta_{n+1}(x) = \sum_{i=1}^{\eta_n(x+1)} \sum_{j=1}^{Z_i^n(x+1)} W_-^{i,j,n,x+1} + \sum_{i=1}^{\eta_n(x-1)} \sum_{j=1}^{Z_i^n(x-1)} W_+^{i,j,n,x-1}$$

with

$$\begin{aligned} &L(\{Z_i^n(x)\}, \{W_+^{i,j,n,x}\}) = [\otimes_{i=1,\dots,\eta_n(x)} F_x] \otimes [\otimes_{i,j \in \mathbb{N},\, n \in N \cup \{0\},\, x \in Z} B(1, \tfrac{1}{2}(1+h))] \\ &W_-^{i,j,n,x} = 1 - W_+^{i,j,n,x} \end{aligned}$$

(L is law, B is the Bernoulli distribution).

0c *Theorems*

We are interested in how $(\eta_n)_{n \in \mathbb{N} \cup \{0\}}$ behaves for fixed F. In order to describe its local and global behavior we introduce the following objects:

$$d_n(F) = E(\eta_n(0) \mid F) \tag{0.5}$$

$$D_n(F, \eta_n) = \lim_{N \to \infty} (2N+1)^{-1} \sum_{x=-N}^{N} \eta_n(x). \tag{0.6}$$

It is straightforward to check that η_n, for each fixed n, is a shift invariant random sequence (when considered also as a random variable in F) satisfying

$$E(\eta_n^2(x)) < \infty$$
$$\lim_{N\to\infty} (2N+1)^{-1} \sum_{x=-N}^{N} \operatorname{Cov}(\eta_n(x), \eta_n(0)) = 0.$$

Here (0.2) and (0.4) are used. Hence η_n is L_2-ergodic and by the ergodic theorem

$$D_n(F, \eta_n) = E(\eta_n(0)) = E(d_n(F)) \quad \text{a.s.} \tag{0.7}$$

We are interested in the long term behavior of $d_n(F)$ and $D_n(F, \eta_n)$. In particular, we want to investigate their dependence on the drift h while keeping the other parameters q, b_1, b_2 and γ fixed.

To formulate our main results, contained in Theorems 1 and 2 below, we need to define the following symbols the role and interpretation of which will become clear along the way:

$$M_\theta = \{\nu \in \wp(\mathbb{N}^2) : \sum_j \nu(i,j) = \sum_j \nu(j,i) \text{ for all } i \in \mathbb{N}, \langle a, \nu\rangle = \theta^{-1}\} \qquad (\theta \in [0,1])$$

$$f(i) = \log(q[b_1]^i + (1-q)[b_2]^i) \qquad (i \in \mathbb{N})$$

$$I_\theta(\nu) = \sum_{i,j} \nu(i,j) \log\left(\frac{\nu(i,j)}{\bar{\nu}(i)\, P_\theta(i,j)}\right) \qquad (\theta \in (-1,1), \nu \in \mathscr{P}(\mathbb{N}^2))$$

$$I_h(\theta) = \frac{1}{2}(1+\theta)\log\left(\frac{1+\theta}{1+h}\right) + \frac{1}{2}(1-\theta)\log\left(\frac{1-\theta}{1-h}\right) \qquad (\theta \in [-1,1], h \in (-1,1))$$

where $\wp(\mathbb{N}^2)$ denotes the set of probability measures on $\mathbb{N}^2$, $\langle \cdot, \cdot \rangle$ denotes inner product, $a(i,j) = i+j-1$, $\bar{\nu}(i) = \sum_j \nu(i,j)$, and

$$P_\theta(i,j) = \binom{i+j-2}{i-1} [\tfrac{1}{2}(1+\theta)]^i [\tfrac{1}{2}(1-\theta)]^{j-1}. \qquad (\theta \in (-1,1), i, j \in \mathbb{N})$$

First we consider the global population density $D_n(F, \eta_n)$ and identify its exponential growth rate in terms of a variational formula.

Theorem 1 (global growth rate): *Under conditions* (0.1–4)

$$\lim_{n\to\infty} \frac{1}{n} \log D_n(F, \eta_n) = \rho(h) \quad \text{a.s.} \tag{0.8}$$

For $h \in (0,1)$

$$\rho(h) = \sup_{\theta \in (0,1)} \sup_{\nu \in M_\theta} (\theta[\langle f \circ a, \nu\rangle - I_\theta(\nu)] - I_h(\theta)). \tag{0.9}$$

For $h = 0$ *and* $h = 1$

$$\rho(0) = \log(\max\{b_1, b_2\})$$
$$\rho(1) = \log(q b_1 + (1-q) b_2). \tag{0.10}$$

Corollary 1 (global localization vs. delocalization and survival vs. extinction):

(1) *$\rho(\cdot)$ is continuous;*

(2) *if $q \in (0,1)$ and $b_1 \neq b_2$, then there exists $h_1 \in (0,1)$ such that*

$$\rho(h) = \log[(1-h^2)^{1/2} \max\{b_1, b_2\}] \quad \text{if } h \leqq h_1$$
$$\rho(h) > \log[(1-h^2)^{1/2} \max\{b_1, b_2\}] \quad \text{if } h > h_1;$$

(3) *if $q \in (0,1)$ and $\max\{b_1, b_2\} > 1 > q b_1 + (1-q) b_2$, then there exists $h_2 \in (0,1)$ where $\rho(\cdot)$ changes sign.*

We shall now give the somewhat informal interpretation of the above results. For a more formal discussion see Baillon et al. [0] Sect. 3.

The variational formula (0.9) tells us that a *selection mechanism* is at work. If in a *homogeneous* branching random walk we randomly pick a particle at time n (by randomly selecting from the population in a large box), say the position of this particle is x, then its ancestor at time 0 was located at site $\approx x - hn$. That is, if with each particle we associate its *path of descent* running backwards in time, then for the particle we randomly pick this path will look like a typical n-step path of the reversed random walk with kernel p_{-h} (see (0.3)) starting at x. However, in the *inhomogeneous* case this is quite different. If $b_1 > b_2$, then paths which happen to spend a lot of time on b_1-sites and little time on b_2-sites create a lot of offspring and therefore contribute substantially to the population at time n. Consequently, for large n most of the population consists of particles whose path of descent *slows down* on b_1-sites and *speeds up* on b_2-sites. The quantitative statement is contained in (0.9): if for the particle we pick randomly at time n at site x we write the position of its ancestor at time 0 as $x - \theta_n n$, then the law of θ_n concentrates on the value θ^* where (0.9) realizes its first supremum. This can be phrased by saying that particles which assume *effective drift* θ^* optimize their progeny. In general θ^* will be different from h.

But there is not only a selection of paths of descent according to their effective drift. Also important is the frequency at which sites are visited along the way. Namely, we shall see that the sequence of local times $\{l(x)\}$ for the path is given by $l(x) = m(x-1) + m(x) - 1$ with $\{m(x)\}$ a Markov sequence (see Sect. 3.a). Consequently, this frequency is controlled by the empirical pair distribution v_n of $\{m(x)\}$ along the path. Equation (0.9) says that the law of v_n concentrates on the measure v^* where the second supremum is realized. We call v^* the *effective empirical pair distribution*. Like θ^* it optimizes the progeny. In general v^* will be different from the equilibrium empirical pair distribution of $\{m(x)\}$ under the kernel $P_h(i,j)$ or $P_{\theta*}(i,j)$.

In fact, the law of $\{m(x)\}$ for the typical path of descent converges to the stationary Markov chain with transition kernel $v^*(i,j)/\bar{v}^*(i)$.

We can now explain the structure of (0.9). If a path of descent has the property that $\theta_n \to \theta$ and $v_n \to v$, then its probability decays at a rate $I_h(\theta) + \theta I_\theta(v)$ per step while it produces offspring at a rate $\theta \sum_{i,j} f(i+j-1)\, v(i,j)$. Here $I_h(\theta)$ is the rate function for the drift and $I_\theta(v)$ is the rate function for the empirical pair distribution of $\{m(x)\}$ along the path under the kernel $P_\theta(i,j)$. Note that θ appears twice as a factor because it determines the range of the path. To match θ and ν we must have $\sum_{i,j}(i+j-1)\,\nu(i,j)=\theta^{-1}$ in M_θ. Hence the *optimal strategy* of the path is to choose $\theta=\theta^*$ and $\nu=\nu^*$, as this maximizes its contribution to the progeny. In other words, the population will predominantly consist of those particles whose path of descent has this limiting behaviour.

A further interesting aspect of (0.9) is that there is a critical value h_1 such that for $h<h_1$ the first supremum is attained at $\theta^*=0$, so that the typical path of descent moves only $o(n)$. This phenomenon we call *localization*. For $h>h_1$, on the other hand, $\theta^*>0$ and the path moves on scale n, which we call *delocalization*. This is also why $\rho(h)$ takes on such a simple analytic form for $h\leqq h_1$ and at the critical value h_1 has a non-analyticity. A related effect was found by Eisele and Lang [8] for the Wiener Sausage with drift.

There is a second critical value h_2 where the process changes from *survival* to *extinction*. This can only happen when the random environment has both super and subcritical offspring distributions.

Now we turn to the local population density $d_n(F)$, which is more difficult because it depends on the environment F. We need the following symbols:

$$\begin{aligned}
&i=(i_1,i_2) \qquad j=(j_1,j_2)\\
&i_1,j_1\in\mathbb{N} \qquad i_2,j_2\in\{b_1,b_2\}\\
&\hat f(i,j)=\log j_2\\
&\hat a(i,j)=i_1+j_1-1\\
&\beta(b_1)=q \qquad \beta(b_2)=1-q
\end{aligned}$$

and

$$M_{\theta,\beta\otimes\beta}=\{\nu\in\wp((\mathbb{N}\otimes\{b_1,b_2\})^2):\sum_j \nu(i,j)=\sum_j \nu(j,i)$$

$$\text{for all } i\in\mathbb{N}, \langle\hat a,\nu\rangle=\theta^{-1}, \tilde\nu=\beta\otimes\beta\}$$

$$I_{\theta,\beta^{\mathbb{Z}}}(\nu)=-\sum_{i,j}\nu(i,j)\log(P_\theta(i_1,j_1)\,\beta(j_2))-\sup_{Q\in A_{\nu,\beta^{\mathbb{Z}}}}(h(Q))$$

$$A_{\nu,\beta^{\mathbb{Z}}}=\{Q\in\wp((N\otimes\{b_1,b_2\})^{\mathbb{Z}}): Q \text{ shift invariant}, \pi^2 Q=\nu, \tilde Q=\beta^{\mathbb{Z}}\}$$

$$h(Q)=\text{Kolmogorov-Sinai entropy of } Q.$$

(Ellis [9] p. 24) where $\tilde Q$ and $\tilde\nu$ are the projections of Q and ν on the medium coordinates, π^2 is taking the two-dimensional marginal of Q.

The exponential growth rate of $d_n(F)$ can be identified in terms of a variational formula. The (limiting) growth rate is F-a.s. independent of the realization of the medium.

Theorem 2 (local growth rate): *Under conditions* (0.1–4)

$$\lim_{n\to\infty}\frac{1}{n}\log d_n(F)=\lambda(h) \qquad F\text{-a.s.} \tag{0.11}$$

For $h\in(0,1)$

$$\lambda(h)=\sup_{\theta\in(0,1)}\sup_{\nu\in M_{\theta,\beta\otimes\beta}}(\theta[\langle\hat f\,\hat a,\nu\rangle-I_{\theta,\beta^{\mathbb{Z}}}(\nu)]-I_h(\theta)). \tag{0.12}$$

For $h=0$ *and* $h=1$

$$\lambda(0)=\log(\max\{b_1, b_2\}) \tag{0.13}$$
$$\lambda(1)=q\log b_1+(1-q)\log b_2.$$

Corollary 2 (local localization vs. delocalization and survival vs. extinction):
(1) $\lambda(\cdot)$ *is continuous*;
(2) *if* $q\in(0,1)$ *and* $b_1 \neq b_2$, *then there exists* $h_3\in(0,1)$ *such that*

$$\lambda(h)=\log[(1-h^2)^{1/2}\max\{b_1, b_2\}] \quad \text{if } h\leqq h_3$$
$$\lambda(h)>\log[(1-h^2)^{1/2}\max\{b_1, b_2\}] \quad \text{if } h>h_3,$$

(3) *if* $q\in(0,1)$ *and* $\log(\max\{b_1, b_2\})>0>q\log b_1+(1-q)\log b_2$, *then there exists* $h_4\in(0,1)$ *where* $\lambda(\cdot)$ *changes sign.*

The same informal interpretations apply as for Theorem 1, with the difference that this time the particle we pick randomly at time n must be chosen from the population at 0 (and not from a large box). Again there is a selection of paths of descent; (0.12) tells us that there is again an effective drift θ^* but this time the effective empirical pair distribution v^* is vector-valued: it is the optimal empirical pair distribution of the Markov sequence (underlying the local times) and the random environment combined, i.e. $\{m(x), b_x\}$. It determines the optimal asymptotic frequencies at which the b_1-sites and the b_2-sites are visited in the fixed environment F and is the 2-dimensional marginal of the optimizing process Q^* for $\{m(x), b_x\}$.

Contrary to the global case, in the local case the law of $\{m(x), b_x\}$ for the typical path of descent converges to the stationary process Q^* that optimizes $I_{\theta,\beta^{\mathbb{Z}}}(v^*)$ and will in general not be Markov. An interesting question that comes up here is: What is the behavior of the optimal path of descent like? Could Q^* be some random walk in random environment? If so, can one calculate its transition probabilities and how do these relate to θ^*, v^*?

In order to be able to speak of a "phase transition" when the system moves locally from survival to extinction one should really also establish that for every h with $\lambda(h)>0$ a *law of large numbers* holds:

$$L(\eta_n(0)/d_n(F)\,|\,F)\to\delta_1 \quad F\text{-a.s. } (n\to\infty). \tag{0.14}$$

This would require techniques which are rather different from the ones used in this paper. We shall discuss this question elsewhere.

A remarkable feature of Theorems 1 and 2 is that apparently $d_n(F)$ and $D_n(F, \eta_n)$ are controlled by different forces.

Corollary 3 (clustering):

(1) $\rho(h)\geqq\lambda(h)$ *for all* $h\in[0,1]$;

(2) $h_1\leqq h_3$ *and* $h_4\leqq h_2$;

(3) *if* $q\in(0,1)$ *and* $b_1\neq b_2$, *then:*
$\rho(h)=\lambda(h)$ *for* $h\leqq h_1$
$\rho(h)>\lambda(h)$ *for* h *close to* 1.

We shall see that the global growth rate is the supremum of local growth rates over random environments that are stationary processes. If both the local and the global optimal path of descent delocalize (i.e. if $h > h_3$), then this is to be interpreted as saying that the global population is mostly made up of particles whose path of descent has moved through a part of the space where the random environment looks like the optimizing environment $\tilde{Q}^*$ rather than the i.i.d. environment with distribution β^Z. If $\tilde{Q}^* \neq \beta^{\mathbb{Z}}$ then this says that the global population density is carried by a thin subset of the space and $\rho(h) > \lambda(h)$. If, on the other hand, both the local and the global optimal path of descent localize (i.e. if $h < h_1$), then the variational problem is degenerate and the paths of descent do not properly sample the medium). Almost all of the time is spent on sites where the growth is maximal, in which case $\rho(h) = \lambda(h) = \log(\{b_1, b_2\}) - I_h(0)$. In the mixed situation (i.e. $h_1 < h < h_3$) we know from Corollary 1 and 2 that $\rho(h) > \lambda(h)$.

0d Extensions

Instead of (0.1) and (0.2) all that is really required to assume about the random environment F is that

(0.15)
$$0 < K^{-1} \leqq \sum_{n=0}^{\infty} n^2 F_x(n) \leqq K < \infty \quad F\text{-a.s.}$$

If we define

(0.16)
$$\sum_{n=0}^{\infty} n F_x(n) = b_x$$
$$\beta = L(b_x),$$

then Theorem 1 remains true with f in (0.9) replaced by

(0.17)
$$f(i) = \log \int b^i \beta(\mathrm{db})$$

and with (0.10) replaced by

(0.18)
$$\rho(0) = \log M$$
$$\rho(1) = f(1)$$

where

(0.19)
$$M = \text{maximal value in support of } \beta.$$

Corollary 1 generalizes accordingly.

The extension of Theorem 2 suggests itself: replace β by (0.16) and $\{b_1, b_2\}$ by supp β and instead of (0.13) the following holds

(0.20)
$$\lambda(0) = \log M$$
$$\lambda(1) = \int \log b\, \beta(\mathrm{d}b).$$

Corollaries 2 and 3 generalize accordingly.

To extend the notion of entropy when supp β is not countable, see e.g. Ellis [9] Theorem A. 9.9. For Theorem 2 we shall assume that β has finite entropy with respect to some appropriate reference measure.

0e Remarks

Corollaries 1, 2 and 3 contain a lot of information about the phase diagram. However, there are still some vital parts missing in order to fully corroborate the qualitative picture drawn in Fig. 1. We conjecture that the following additional properties hold (when β has positive variance):

(1) $\rho(\cdot)$ and $\lambda(\cdot)$ are strictly decreasing on $[0, 1]$;
(2) $\rho(h) > \lambda(h)$ for all $h > h_1$;
(3) $\theta^*(h) < h$ for all $0 < h < 1$ both locally and globally;
(4) $h_1 < h_3$ and $h_4 < h_2$.

The investigation of these properties requires functional analytic techniques of some depth. We adress this question in Baillon et al. [1]. In Sect. 6 we shall see why (1–4) are quite plausible.

In Baillon et al. [0] and in Greven and den Hollander [15] we study the simpler version of our model where the random walk kernel in (0.3) is replaced by $p_h(x, y) = 1 - h$ for $y = x$ and $p_h(x, y) = h$ for $y = x + 1$. For this situation the analysis simplifies considerably, we end up with variational formulas that can be solved explicitly, and the properties (1)–(4) are verified.

Clearly, (1) implies that h_2 and h_4 are unique (provided they exist). An interesting consequence of (2) and (4) is the existence of an *intermediate phase* of *extreme clustering*:

$$\rho(h) > 0 > \lambda(h) \qquad \text{for } h_4 < h < h_2.$$

This means that an overwhelming part of the population lives on a random set thinning out as time proceeds but carrying fast growing clusters of particles. Naturally, this raises the question how one can describe this set and what the clusters look like.

Finally, we have seen that the optimal path of descent slows down on sites where b_x is large and speeds up on sites where b_x is small; (3) says that the net effect of the random environment is to slow the path down.

The rest of this paper is devoted to the proofs. In Sect. 1 we first outline the general scheme of proof and isolate the mathematical problems whose treatment is the core of this work.

1 General scheme of proof and key propositions

This section consists of three parts in which we formulate four key propositions (Propositions 2–5) and explain how these imply Theorems 1 and 2. These propositions will then be proved in Sects. 2–5. Corollaries 1–3 will be proved in Sect. 6, where we analyze the variational formulas.

1 a Duality with random walk in random scenery

Let $(S_k)_{k\in\mathbb{N}\cup\{0\}}$ be random walk with transition kernel $\hat{p}_h(x,y)=p_h(y,x)$ starting at 0, let $\hat{E}_h$ denote expectation with respect to it, and define the local times at time n

$$l_n(x)=\sum_{k=1}^{n}\chi\{S_k=x\}\qquad (x\in\mathbb{Z},\, n\in\mathbb{N}). \tag{1.1}$$

Proposition 0 *With* γ, b_x *and* f *as defined in* (0.4), (0.16) *and* (0.17)

$$d_n(F)=\gamma\hat{E}_h(\exp[\sum_{x\in\mathbb{Z}} l_n(x)\log(b_x)]) \tag{1.2}$$

$$D_n(F,\eta_n)=\gamma\hat{E}_h(\exp[\sum_{x\in\mathbb{Z}} f(l_n(x))])\quad \text{a.s.} \tag{1.3}$$

Proof. Introduce an independent collection of single ancestor branching random walks

$$(\eta_n^{y,k})_{n\in N}\qquad (y\in\mathbb{Z},\, k\in\mathbb{N}).$$

Here for each k, $(\eta_n^{y,k})_{n\in\mathbf{N}\cup\{0\}}$ is our process starting from one particle at site y and no particles elsewhere. If we add to this collection an independent realization of the initial configuration η_0, then we can define a version of our process as follows:

$$\eta_n(x)=\sum_{y\in\mathbb{Z}}\sum_{k=1}^{\eta_0(y)}\eta_n^{y,k}(x).$$

This gives

$$E(\eta_n(x)\mid F)=\gamma\sum_{y\in\mathbb{Z}} E(\eta_n^{y,1}(x)\mid F). \tag{1.4}$$

Next use the fact that at each step each particle at site x independently creates offspring of expected size b_x, to calculate

$$\begin{aligned} E(\eta_n^{y,1}(x)\mid F) &= \sum_{z_1,\ldots,z_{n-1}}[b_y\, p_h(y,z_1)]\ldots[b_{z_{n-1}}\, p_h(z_{n-1},x)]\\ &= \sum_{z_1,\ldots,z_{n-1}}[\hat{p}_h(x,z_{n-1})\, b_{z_{n-1}}]\ldots[\hat{p}_h(z_1,y)\, b_y]\\ &= \hat{E}_h\left(\exp\left[\sum_{k=1}^{n}\log(b_{x+S_k})\right]\chi\{S_n=y-x\}\right). \end{aligned} \tag{1.5}$$

The last equality uses the shift invariance of $\hat{p}_h$. Combination of (0.5), (1.1), (1.4) and (1.5) gives (1.2). Now (1.3) follows by averaging over F in (1.2) using (0.7), (0.17), the i.i.d. property of the random environment, and Fubini' theorem. □

1b Dependence on the drift h

From now on we drop the superscript $\hat{}$ because instead of $\hat{p}_h = p_{-h}$ we may as well work with p_h and assume $h \in [0,1]$. Let P_h and E_h denote probability and expectation for the random walk with kernel p_h. Our first observation is that the quantities in (1.2) and (1.3) are of the form

$$E_h(\exp[\sum_x f_x(l_n(x))]) \tag{1.6}$$

with, respectively,

$$f_x(i) = \begin{cases} i \log(b_x) \\ f(i). \end{cases}$$

Write (1.6) as

$$\int_{\theta \in [-1,1)} E_h(\exp[\sum_x f_x(l_n(x))] \mid S_n = [\theta n]) \, P_h(S_n \in \mathrm{d}[\theta n]) + \text{integrand at } \theta = 1.$$

Since all n-step paths have equal probability when conditioned on $S_n = [\theta n]$ (with $[x]$ denoting the largest integer $\leqq x$), we have

$$E_h(\exp[\sum_x f_x(l_n(x))] \mid S_n = [\theta n]) \qquad \text{is } \textit{constant} \text{ in } h.$$

Therefore, replacing h by θ we get for (1.6)

$$\int_{\theta \in [-1,1]} E_\theta(\exp[\sum_x f_x(l_n(x))] \mid S_n = [\theta n]) \, P_h(S_n \in \mathrm{d}[\theta n]) + \text{integrand at } \theta = 1. \tag{1.7}$$

The significance of this expression lies in the fact that the drift parameter h *only appears in the integrating measure.* This allows us to isolate the drift dependence:

Proposition 1 *Suppose that there exists $J: [-1,1] \to \mathbb{R}$ bounded and continuous such that*

$$\lim_{n \to \infty} \frac{1}{n} \log E_\theta(\exp[\sum_x f_x(l_n(x))] \, \chi\{S_n = [\theta n]\}) = J(\theta) \qquad F\text{-a.s.} \tag{1.8}$$

and suppose that the same limit is obtained along any sequence $\theta_n \to \theta$. Then for $h \in [0,1]$, and with $I_h(\theta)$ as defined prior to Theorem 1,

$$\lim_{n \to \infty} \frac{1}{n} \log E_h(\exp[\sum_x f_x(l_n(x))]) = \sup_{\theta \in [-1,1]} (\mathrm{J}(\theta) - \mathrm{I}_h(\theta)) \qquad F\text{-a.s.} \tag{1.9}$$

Proof. It is well known that $P_h(S_n/n \in \mathrm{d}\theta)$ is a large deviation family with rate function $I_h(\theta)$ (Ellis [9] p. 11). Since $I_h(\theta)$ is continuous in θ, it follows that the same is true for $P_h(S_n \in d[\theta n])$. Hence (1.9) follows by applying Varadhan's Theorem (see [20], Theorem 3.4) to (1.7), using (1.8). We also need that

$$P_\theta^{-1}(S_n=[\theta n])=O(n^{1/2})=\exp(o(n)) \qquad \text{uniformly in } \theta$$

to turn the condition in (1.7) into an indicator in (1.8). □

If $J^\lambda(\cdot)$ and $J^\rho(\cdot)$ denote the J-functions in (1.8) corresponding to the two choices of f_x in (1.6), then (1.9) tells us that

$$\lambda(h)=\sup_{\theta\in[-1,1]}(J^\lambda(\theta)-I_h(\theta)) \tag{1.10}$$
$$\rho(h)=\sup_{\theta\in[-1,1]}(J^\rho(\theta)-I_h(\theta)).$$

Thus, to prove Theorems 1 and 2 what we have to do is verify the assumptions in Proposition 1 identify $J^\lambda(\theta)$ and $J^\rho(\theta)$, and argue that the supremum may be restricted to $\theta\in(0,1)$. Equation (1.10) will then give us (0.9) and (0.12). The continuity and boundedness of $J(\cdot)$ will be settled in Sect. 6. At the end of Sects. 3 and 5 we check that for $n\to\infty$

$$\frac{1}{n}\log E_{\theta_n}(\exp[\sum_x f_x(l_n(x))]\,\chi\{S_n=[\theta_n n]\})\to J(\theta) \qquad \text{for every } \theta_n\to\theta. \tag{1.11}$$

1c Proof of Theorem 1 and 2

The following two propositions combined prove existence of and identify $J^\rho(\theta)$. Let

$$l(x)=\sum_{k=0}^{\infty}\chi\{S_k=x\}=\delta_0(x)+\lim_{n\to\infty} l_n(x) \qquad (x\in\mathbb{Z}). \tag{1.12}$$

Proposition 2 *For every* $\theta\in(0,1)$

$$E_\theta(\exp[\sum_x f(l_n(x))]\,\chi\{S_n=[\theta n]\}) \tag{1.13}$$
$$=\exp(o(n))\,E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]} f(l(x))\right]\chi\left\{\sum_{x=0}^{[\theta n]} l(x)=n\right\}\chi\{l(0)=l([\theta n])=1\}\right).$$

Proposition 3 *For every* $\theta\in(0,1)$

$$\lim_{n\to\infty}\frac{1}{n}\log E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]} f(l(x))\right]\chi\left\{\sum_{x=0}^{[\theta n]} l(x)=n\right\}\chi\{l(0)=l([\theta n])=1\}\right) \tag{1.14}$$
$$=\theta\sup_{\nu\in M_\theta}[\langle f\circ a,\nu\rangle-I_\theta(\nu)].$$

The proof of these two propositions will be given in Sects. 2 and 3, respectively.

The following two propositions combined prove existence of and identify $J^\lambda(\theta)$.

Proposition 4 *For every* $\theta \in (0, 1)$

(1.15) $E_\theta(\exp[\sum_x l_n(x)\log(b_x)]\,\chi\{S_n=[\theta n]\})$

$$=\exp(o(n))\,E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]} l(x)\log(b_x)\right]\chi\left\{\sum_{x=0}^{[\theta n]} l(x)=n\right\}\chi\{l(0)=l([\theta n])=1\}\right),\quad \text{F-a.s.}$$

Proposition 5 *For every* $\theta \in (0, 1)$

$$(1.16)\quad \lim_{n\to\infty}\frac{1}{n}\log E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]} l(x)\log(b_x)\right]\chi\left\{\sum_{x=0}^{[\theta n]} l(x)=n\right\}\chi\{l(0)=l([\theta n])=1\}\right)$$
$$=\theta \sup_{v\in M_{\theta,\beta\otimes\beta}} [\langle \hat f \hat a, v\rangle - I_{\theta,\beta^{\mathbb{Z}}}(v)]\qquad F\text{-a.s.}$$

The proof is in Sects. 4 and 5, respectively.

We remark that the extra indicator in (1.13–16) requiring that $l(0)=l([\theta n])=1$ is important and its origin will become clear from the proofs.

The regularity suppositions in Proposition 1 will be verified at the end of Sects. 3 and 5.

Combination of Propositions 2–5 with (1.10) completes the proof of (0.9) and (0.12) after the following observations concerning the range $\theta\in(0,1)$ in the variational formulas.

Since $J^\lambda(\theta)$ and $J^\rho(\theta)$ are symmetric in θ and since $I_h(\theta)\leqq I_h(-\theta)$ for all $\theta\in[-1,1]$, $h\geqq 0$, the supremum in (1.10) may be restricted to $\theta\in[0,1]$. The boundary cases $\theta=0$ and $\theta=1$ are degenerate. Indeed, if $\theta=1$ then $l_n(x)=1$ for $0<x\leqq n$ and zero elsewhere, and substitution into (1.8) immediately gives

$$J^\lambda(1)=\int \log b\,\beta(\mathrm{d}b)$$
$$J^\rho(1)=f(1).$$

For $\theta=0$, on the other hand, the random walk is symmetric and it was shown by Greven [13] that

$$J^\lambda(0)=J^\rho(0)=\log M.$$

(The idea behind this equality is that for symmetric random walk the growth rate is maximal because the probability for the walk to stay in a large box consisting of sites where b_x is maximal decays at a rate which tends to zero as the box becomes large.) Thus we know the cases $\theta=0$ and $\theta=1$, which is why for the proof of Propositions 2–5 we need only consider $\theta\in(0,1)$. Now, in Sect. 6 we shall see that $J^\lambda(\theta)$ and $J^\rho(\theta)$ are continuous, in particular at $\theta=0$ and $\theta=1$. Since $I_h(\theta)$ is continuous in θ, this explains why in (0.9) and (0.12) the supremum over θ may be restricted to (0, 1).

Finally, for all $\theta\in(0,1]$ we have from (0.18)

$$J^\lambda(\theta)\leqq J^\lambda(0)$$
$$J^\rho(\theta)\leqq J^\rho(0)$$

since $f_x(i)\leqq i\log M$ and $\sum_x l_n(x)=n$. So now also (0.18) and (0.20) follow, because $I_0(\theta)>I_0(0)=0$ for $\theta\neq 0$, $I_1(1)=0$ and $I_1(\theta)=\infty$ for $\theta\neq 1$. $\square$

2 Proof of Proposition 2

Equation (1.13) states that, up to a subexponential factor, its r.h.s. and l.h.s. are determined by those paths that stay between 0 and $[\theta n]$ until time n, cross $[\theta n]$ at time n, and then stay to the right of $[\theta n]$ forever. That is, on a scale of order n the expectations apparently are determined by those paths that move out to infinity at speed θ equal to the drift of the underlying random walk kernel. In other words, the indicators in (1.13) force the path to assume *effectively drift θ at any place and at any time.* On a scale $o(n)$ there will of course be fluctuations, and we must build these in the proof without affecting θ. The proof of (1.13) consists of mappings of paths and of combinatorial estimates showing that indeed paths not having effective drift θ contribute negligibly.

This section consists of two parts in which we prove upper and lower bounds, the upper bound being the hard one. The following properties of the function f in (0.17) hold in general and are all that will be needed for the proof:

$$\begin{aligned} &f(i) = i \log M - g(i) \\ &g(0) = 0 \\ &g(\cdot) \text{ is non-decreasing and concave} \\ &g(i) = o(i) \text{ as } i \to \infty. \end{aligned} \tag{2.1}$$

Before we start the proof let us agree on some notation. It will be expedient to extend the random walk to negative times by running an copy of the reversed random walk with opposite drift $-\theta$ from 0 and conditioning it on never returning to 0. In this way we get a two-sided path, denoted by

$$S = (S_n)_{n \in \mathbb{Z}} \qquad (S_0 = 0),$$

which is the random walk with drift θ starting from $-\infty$ at time $-\infty$ and conditioned on first hitting 0 at time 0. Define

$$l_n(x, S) = \sum_{k=1}^{n} \chi\{S_k = x\} \qquad (x \in \mathbb{Z}, n \in \mathbb{N})$$

$$l(x, S) = \sum_{k=-\infty}^{\infty} \chi\{S_k = x\} \qquad (x \in \mathbb{Z}).$$

Note that $l_n(x, S)$ is the same as $l_n(x)$ in (1.1), while $l(x, S)$ is the same as $l(x)$ in (1.12) for $x \geqq 0$. It is straightforward to check that $\{l(x, S)\}$ is shift-invariant on $\mathbb{Z}$. (Use the strong Markov property of random walk.) By E_θ we shall denote expectation w.r.t. S.

Furthermore, we pick a function $\delta: \mathbb{N} \to \mathbb{N}$ such that

$$\begin{aligned} &\delta(n) \to \infty \\ &\delta(n)\, g(n) = o(n) \qquad (n \to \infty) \end{aligned} \tag{2.2}$$

and n will be taken large enough so that $\delta(n)<[\theta n]$. We abbreviate

$$I_n=[0,[\theta n]]$$
$$J_n=[-\delta(n),[\theta n]+\delta(n)].$$

In the sequel the symbols $\cong$, $\gtrsim$, and $\lesssim$ will be used to denote (in)equality up to a subexponential factor $\exp(o(n))$.

2a *Upper bound*

The object of this section is to prove that

$$\text{(2.3)} \qquad \text{l.h.s.}(1.13)\gtrsim \text{r.h.s.}(1.13).$$

First use (2.1) and write

$$\text{(2.4)} \qquad \text{l.h.s.}(1.13)=M^n E_\theta(\exp[-G_n(S)]\,\chi\{S\in A_n^1\})$$

with

$$\text{(2.5)} \qquad G_n(S)=\sum_{x=-\infty}^{\infty} g(l_n(x,S))$$
$$A_n^1=\{S\colon S_n=[\theta n]\}.$$

The first lemma shows that the walk does not want to leave $\operatorname{int}(J_n)$ between time 0 and n, the idea being that anything the walk can do outside I_n it can also do inside I_n. Let

$$A_n^2=A_n^1\cap\{S\colon S_i\in\operatorname{int}(J_n)\ \text{ for } 0<i<n\}.$$

Lemma 1 $E_\theta(\exp[-G_n(S)]\,\chi\{S\in A_n^1\})\cong E_\theta(\exp[-G_n(S)]\,\chi\{S\in A_n^2\})$.

Proof. It suffices to show that

$$E_\theta(\exp[-G_n(S)]\,\chi\{S\in A_n^1\backslash A_n^2\})\lesssim E_\theta(\exp[-G_n(S)]\,\chi\{S\in A_n^2\}).$$

We shall do this by comparing paths in $A_n^1\backslash A_n^2$ and A_n^2, for which we shall need the following definition:

(i) a *small excursion* to the left of x is a path beginning and ending at x and staying inside $(x-\delta(n),x)$;

(ii) a *large excursion* to the left of x is a path beginning and ending at x, staying inside $(-\infty,x)$ but not inside $(x-\delta(n),x)$;

similarly for small and large excursions to the right of x. Thus, small excursions outwards from $\partial(I_n)$ stay inside $\operatorname{int}(J_n)$, but large excursions do not.

Next, let

$$T_n\colon\ A_n^1\backslash A_n^2\to A_n^2$$

be the following map *acting on the path S between time* 0 *and n* (and leaving the rest invariant):

First the large excursions to the left of 0 are reflected around 0.
Then the large excursions to the right of $[\theta n]$ are reflected around $[\theta n]$, including any new large excursions that may have arisen from the previous reflection. This procedure is repeated until the image path lies entirely inside $\mathrm{int}(J_n)$.

In other words: T_n maps a path in $A_n^1 \setminus A_n^2$ onto an image path in A_n^2 by alternately folding large excursions outwards from $\partial(I_n)$ around $\partial(I_n)$. Only the large excursions occurring between time 0 and n are reflected around $\partial(I_n)$ until they lie inside $\mathrm{int}(I_n)$; the small excursions are untouched, as well as the large excursions not occurring between time 0 and n.

We shall need three properties of T_n:

(2.6) T_n preserves probability: any path and its image under T_n have the same probability;

(2.7) $G_n(T_n(S)) - G_n(S) \leqq o(n)$ uniformly for $S \in A_n^1 \setminus A_n^2$.

(2.8) $|\{S \in A_n^1 \setminus A_n^2 : T_n(S) = S'\}| = \exp(o(n))$ uniformly for $S' \in A_n^2$.

Proof of (2.7). Split $l_n(x, S)$ for $x \notin I_n$ into contributions coming from small respectively large excursions outwards from $\partial(I_n)$:

$$l_n(x, S) = l_n^<(x, S) + l_n^>(x, S) \qquad (x \notin I_n).$$

Next split $l_n^>(x, S)$ further into contributions coming from parts of the path that under T_n are mapped inside $\mathrm{int}(I_n)$ respectively $\mathrm{int}(J_n) \setminus I_n$:

$$l_n^>(x, S) = l_n^{>,+}(x, S) + l_n^{>,-}(x, S) \qquad (x \notin I_n).$$

Now use (2.1), by which $g(i) + g(j) \geqq g(i+j)$ for all i and j, to obtain

$$\sum_{x \in I_n} g(l_n(x, S)) + \sum_{x \notin I_n} g(l_n^{>,+}(x, S)) \geqq \sum_{x \in I_n} g(l_n(x, T_n(S))).$$

Here the crucial point is that T_n stacks local times $l_n^{>,+}(x, S)$ at sites $x \notin I_n$ on top of local times $l_n(x, S)$ at sites $x \in I_n$. It is important to realize that for each $x \notin I_n$ all of $l_n^{>,+}(x, S)$ is mapped onto a single site inside I_n, so that the inequality follows by induction on the number of reflections. Thus, for $S \in A_n^1 \setminus A_n^2$,

$$G_n(T_n(S)) - G_n(S) \leqq \sum_{x \notin I_n} [g(l_n^{>,+}(x, S)) - g(l_n(x, S)) + g(l_n(x, T_n(S)))].$$

Since g is increasing and since $l_n(x, T_n(S)) = 0$ for $x \notin \mathrm{int}(J_n)$ this gives via (2.2)

$$G_n(T_n(S)) - G_n(S) \leqq \sum_{x \in \mathrm{int}(J_n) \setminus I_n} g(l_n(x, T_n(S))) \leqq 2\delta(n)\, g(n) = o(n). \qquad \square$$

Proof of (2.8). Here just note that for each $S \in A_n^1 \setminus A_n^2$ the total number of large excursions that get reflected by T_n is at most $n/2\delta(n)$ (= length of path prior to time n/minimum length of large excursion). Each reflection gives rise to a multiplicity 2: in $T_n(S)$ each large excursion to the right of 0 or to the left of $[\theta n]$ can either occur already in S or can be the reflected image of a large excursion in S to the left of 0 or to the right of $[\theta n]$. Hence the total multiplicity of S is at most $2^{n/2\delta(n)} = \exp(o(n))$. $\square$

Applying first (2.7) and then (2.6) and (2.8), we obtain

$$\begin{aligned} E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^1 \setminus A_n^2\}) &\gtrsim E_\theta(\exp[-G_n(T_n(S))]\,\chi\{S \in A_n^1 \setminus A_n^2\}) \\ &\cong E_\theta(\exp[-G_n(S')]\,\chi\{S' \in A_n^2\}), \end{aligned}$$

which completes the proof of Lemma 1. $\square$

The second lemma shows that, because of the positive drift θ, the walk quickly enters J_n from the left before time 0 and quickly leaves J_n from the right after time n. Let

$$\begin{aligned} A_n^3 = A_n^2 &\cap \{S : S_j = j \text{ for } -\delta(n) \leqq j < 0,\ S_j < -\delta(n) \text{ for } j < -\delta(n)\} \\ &\cap \{S : S_{n+j} = [\theta n] + j \text{ for } 0 < j \leqq \delta(n),\ S_{n+j} > [\theta n] + \delta(n) \text{ for } j > \delta(n)\}. \end{aligned}$$

Lemma 2 $E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^2\}) \cong E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^3\})$.

Proof. Observe that $G_n(S)$ only depends on S between time 0 and n and that the difference between A_n^2 and A_n^3 only involves what S does before and afterwards. Therefore the Markov property gives

$$E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^2\}) = P_\theta^{-1}(A_n^3 \mid A_n^2)\, E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^3\}).$$

But

$$P_\theta(A_n^3 \mid A_n^2) = \theta[\tfrac{1}{2}(1+\theta)]^{2\delta(n)},$$

since a step to the right has probability $\frac{1}{2}(1+\theta)$ and escaping to the right has probability θ. (Recall here the reversed random walk that was used to extend the path S to negative times.) Now use that $\delta(n) = o(n)$. $\square$

At this point we know enough to replace $l_n(x, S)$ by $l(x, S)$ and restrict the sum (2.5) in the exponent of (2.4) to J_n. Therefore define

$$G_n^*(S) = \sum_{x \in J_n} g(l(x, S)). \tag{2.9}$$

Lemma 3 $E_\theta(\exp[-G_n(S)]\,\chi\{S \in A_n^3\}) \cong E_\theta(\exp[-G_n^*(S)]\,\chi\{S \in A_n^3\})$.

Proof. On A_n^3 we have (recall (1.12))

$$\begin{aligned} l_n(x, S) &= l(x, S) - \delta_0(x) && \text{for } x \in I_n \\ l_n(x, S) &= l(x, S) - 1 && \text{for } x \in J_n \setminus I_n \\ l_n(x, S) &= 0 && \text{for } x \notin \operatorname{int}(J_n). \end{aligned}$$

Hence from comparison of (2.5) and (2.9), and since $g(0) = 0$,

$$0 \leqq G_n^*(S) - G_n(S) = \sum_{x \in (J_n \setminus I_n) \cup \{0\}} [g(l(x, S)) - g(l(x, S) - 1)] \leqq (2\delta(n) + 1)\, g(1) = o(n),$$

where the inequality follows from $g(i) - g(i-1) \leqq g(1)$ via (2.1). $\square$

Combining Lemmas 1–3 with (2.4) we arrive at

$$\text{l.h.s.}(1.13) \cong M^n E_\theta(\exp[-G_n^*(S)]\,\chi\{S \in A_n^3\}). \tag{2.10}$$

Finally, note that

$$A_n^3 \subset \{S: \sum_{x \in J_n} l(x, S) = n + 2\delta(n),\ l(-\delta(n), S) = l([\theta n] + \delta(n), S) = 1\}$$

to obtain via (2.1) that

$$\begin{aligned}\text{l.h.s.}(1.13) &\lesssim E_\theta(\exp[\sum_{x \in J_n} f(l(x, S))]\,\chi\{\sum_{x \in J_n} l(x, S) = n + 2\delta(n)\} \\ &\quad \times \chi\{l(-\delta(n), S) = l([\theta n] + \delta(n), S) = 1\}) \\ &= E_\theta\left(\exp\left[\sum_{x=0}^{[\theta n]+2\delta(n)} f(l(x, S))\right]\chi\left\{\sum_{x=0}^{[\theta n]+2\delta(n)} l(x, S) = n + 2\delta(n)\right\}\right. \\ &\quad \left.\times \chi\{l(0, S) = l([\theta n] + 2\delta(n), S) = 1\}\right).\end{aligned} \tag{2.11}$$

The latter equality uses the stationarity of $\{l(x, S)\}$. Now, the r.h.s. of (2.11) almost equals the r.h.s. of (1.13), the only difference being that n and $[\theta n]$ are perturbed by terms of order $\delta(n) = o(n)$. However, from the proof of Proposition 3 in Sect. 3 it will become clear that this is a lower order effect (see the end of Sect. 3.e), i.e.

$$\text{r.h.s.}(2.11) \cong \text{r.h.s.}(1.13). \quad \square \tag{2.12}$$

2b Lower bound

The proof of the opposite inequality

$$\text{l.h.s.}(1.13) \gtrsim \text{r.h.s.}(1.13) \tag{2.13}$$

is trivial. Just note that

$$\{S: \sum_{x \in I_n} l(x, S) = n,\ l(0, S) = l([\theta n], S) = 1\} \subset \{S: S_n = [\theta n]\}.$$

Indeed, since 0 and $[\theta n]$ are hit only once the path spends all its local time inside I_n in one piece: after entering I_n at time 0 it must stay inside I_n during n steps and cross $[\theta n]$ at exactly time n. This, moreover, implies that $l(x, S) = \delta_0(x) + l_n(x, S)$ for $x \in I_n$ and $l_n(x) = 0$ for $x \notin I_n$ (recall (1.12)), so that via (2.1)

$$\sum_{x \in Z} f(l_n(x, S)) \geqq \sum_{x \in I_n} f(l(x, S)) - \log M. \quad \square$$

Combination of (2.3) and (2.13) completes the proof of Proposition 2. $\square$

3 Proof of Proposition 3

3a $l(x)$ as functional of a Markov process

We start by introducing the Markov process driving the local times. Let

(3.1) $$m(x)=\sum_{k=0}^{\infty}\chi\{S_k=x, S_{k+1}=x+1\} \qquad (x\in\mathbb{Z})$$

denote the total number of steps $x\to x+1$ in the random walk. Since the path is continuous and has positive drift, every step $x\to x-1$ must eventually be followed by a step $x-1\to x$ (with probability 1). Hence (recall (1.12)):

(3.2) $$l(x)=m(x-1)+m(x)-1 \qquad (x\geqq 1).$$

This is nice because $m(x)$ has the following properties.

Lemma 4 *The sequence $\{m(x)\}_{x\geqq 0}$ is stationary Markov with transition kernel and invariant probability measure*

(3.3) $$P_\theta(i,j)=\binom{i+j-2}{i-1}[\tfrac{1}{2}(1+\theta)]^i[\tfrac{1}{2}(1-\theta)]^{j-1} \qquad (i,j\geqq 1),$$

(3.4) $$\pi_\theta(i)=\frac{2\theta}{1+\theta}\left(\frac{1-\theta}{1+\theta}\right)^{i-1} \qquad (i\geqq 1).$$

In fact $\{m(x)\}_{x\geqq 0}$ is a branching process with one immigrant and with subcritical offspring distribution:

(3.5) $$m(x+1)=Z_1+\dots+Z_{m(x)}+1$$

(3.6) $$P(Z_i=k)=\tfrac{1}{2}(1+\theta)[\tfrac{1}{2}(1-\theta)]^k \qquad (k\geqq 0).$$

Proof. For every $x\geqq 0$

$$m(x+1)=Z_1+\dots+Z_{m(x)-1}+Z'$$

with

$Z_i=\#$ steps $x+1\to x+2$ between i-th and $(i+1)$-st step $x\to x+1$
$Z'=\#$ steps $x+1\to x+2$ following last step $x\to x+1$.

The Z_i are i.i.d. and count the excursions to the right of $x+1$ until the next step $x\to x+1$; Z' is independent of the Z_i and $Z'-1$ counts the excursions to the right of $x+1$ following the last step $x\to x+1$. Since each of these excursions has probability $\frac{1}{2}(1-\theta)$, it follows that the Z_i and $Z'-1$ have the distribution given in (3.6), and also that (3.5) holds. Now (3.3) is straightforward, and (3.4) follows by checking that $\pi_\theta(i)\,P_\theta(i,j)=\pi_\theta(j)\,P_\theta(j,i)$, so that π_θ is a reversible equilibrium. Finally, the stationarity follows by using (3.1) to check that $P(m(0)=i)=\pi_\theta(i)$, so that the process, tarts in equilibrium. □

In view of (3.2), our process of local times is a simple *two-block functional of a Markov process.* This is nice because it means that at this stage in the proof we can forget about the random walk: Proposition 3 is a *large deviation* problem for a Markov process of a very specific structure. In particular, what we need is a large deviation property for the empirical process of pairs associated with $\{m(x)\}_{x\geqq 0}$, the so-called *level-2 analysis for pairs*, under the sum restriction enforced by the indicator in the l.h.s. of (1.14). Via (3.2) this will give us a large deviation property for the empirical process of $\{l(x)\}_{x\geqq 0}$. This is what we are looking for because the exponent in the l.h.s. of (1.14) can be expressed as a functional of this empirical process.

For Markov processes large deviations have been studied quite extensively in the papers by Donsker and Varadhan [7], Stroock [19], Ney and Nummelin [18]. However, our process has infinite state space N and is not uniformly recurrent, so it does not belong to the class of countable Markov processes for which level-2 large deviation principles have been derived in the literature. Therefore we shall need to do some work to get Proposition 3 going.

The rest of this section consists of four parts. In Sect. 3.b we formulate the right framework by using the Markov property of $\{m(x)\}$ to reduce the problem to i.i.d. random variables. We then point out a number of technical difficulties. In Sects. 3.c and 3.d we do a truncation and perturbation analysis in order to circumvent these difficulties and to prepare for the final large deviation analysis in Sect. 3.e. The analysis in Sects. 3.c–e is nonstandard because of the presence of the indicator in (1.14). On the one hand, this forces us to establish a large deviation principle for a sequence of *conditional* probability measures. On the other hand, the indicator plays an important role in handling the infinite state space.

3b Passing to i.i.d. random variables

For K and L positive integer let us abbreviate

$$\text{(3.7)}\qquad E_\theta(K,L)=E_\theta\left(\exp\left[\sum_{x=0}^{K} f(l(x))\right]\chi\left\{\sum_{x=0}^{K} l(x)=L\right\}\chi\{l(0)=l(K)=1\}\right).$$

In the previous section we have seen that $E_\theta(K,L)$ is the expectation of a functional of a Markov process, $\{m(x)\}$. In this section we shall further simplify $E_\theta(K,L)$ by rewriting it as a new functional of an i.i.d. sequence. This reduction will at first appear rather hopeful in that it will seem to lead us to known territory, but unfortunately this is not the case. We shall then point out what the main obstacles are that have to be removed and indicate how this will be carried through in Sect. 3.c and 3.d.

Substitute (3.2) into (3.7) and write

$$E_\theta(K,L)=E_\theta\left(\exp\left[\sum_{x=0}^{K} f(m(x-1)+m(x)-1)\right]\chi\left\{\sum_{x=0}^{K}[m(x-1)+m(x)-1]=L\right\}\right.$$
$$\left.\times\chi\{m(-1)=m(0)=m(K-1)=m(K)=1\}\right)$$
$$=\exp(f(1))[\tfrac{1}{2}(1+\theta)]\,E_\theta\left(\exp\left[\sum_{x=1}^{K} f(m(x-1)+m(x)-1)\right]\right.$$
$$\left.\times\chi\left\{\sum_{x=1}^{K-2} m(x)=\tfrac{1}{2}(L+K-5)\right\}\chi\{m(0)=m(K-1)=1\}\right).$$

Here we first use the Markov property of $\{m(x)\}$ to get rid of the left boundary term $m(-1)=1$, and then we introduce *periodic boundary conditions* by putting $m(K)=m(0)$ in the last term of the sum in the exponent (this just comes in handy). Next write out the probabilities of $\{m(x)\}_{x=0}^{K}$ by inserting the kernel (3.3), use (3.4) for the initial value $m(0)=1$, and write

$$E_\theta(K,L)=\theta\exp(f(1))\sum_{m\in V(K,L)}\exp[F(m)+P(m)] \tag{3.8}$$

where

$$V(K,L)=\left\{m\in\mathbb{N}^K:\ \sum_{x=1}^{K-2} m(x)=\tfrac{1}{2}(L+K-5),\ m(0)=m(K-1)=1\right\} \tag{3.9}$$

$$F(m)=\sum_{x=1}^{K} f(m(x-1)+m(x)-1) \tag{3.10}$$

$$P(m)=\sum_{x=1}^{K}\log P_\theta(m(x-1),m(x)). \tag{3.11}$$

Our next step is to introduce i.i.d. random variables $X_1,\dots,X_{K-2}$ with common geometric distribution

$$P(X_k=l)=(1-c)\,c^{l-1}\qquad(l\geqq 1)$$

where $c\in(0,1)$ is arbitrary. In terms of these *auxiliary* objects we may write

$$E_\theta(K,L)=\theta\exp(f(1))(1-c)^{-K+2}c^{-(L-K-1)/2}S(K,L) \tag{3.12}$$

where we define

$$S(K,L)=E(\exp[F(X)+P(X)]\,\chi\{X\in V(K,L)\}) \tag{3.13}$$
$$X=(1,X_1,\dots,X_{K-2},1)$$

and E denotes expectation w.r.t. X. Here we use that all X have the same probability $(1-c)^{K-2}c^{\frac{1}{2}(L+K-5)-K-2}$, because $V(K,L)$ fixes both length and sum of

X. Finally, to turn (3.13) into a more standard form we introduce the empirical pair distribution of X

$$\nu_K = K^{-1} \sum_{k=1}^{K} \delta_{(X_{k-1}, X_k)}$$

(recall $X_K = X_0$) and write

Lemma 5 *For every* $\theta \in (0, 1)$

$$S(K, L) = E(\exp\{K[\hat{F}(\nu_K) + \hat{P}(\nu_K)]\} \chi\{\nu_K \in \hat{V}(K, L)\}) \tag{3.14}$$

where

$$\hat{V}(K, L) = \{\nu \in \wp(\mathbb{N}^2): \sum_j \nu(i, j) = \sum_j \nu(j, i) \text{ for all } i \in \mathbb{N}, \tag{3.15}$$
$$\sum_{i,j} (i + j - 1)\, \nu(i, j) = (L-1)/K\}$$

$$\hat{F}(\nu) = \sum_{i,j} \nu(i, j)\, f(i + j - 1) \tag{3.16}$$

$$\hat{P}(\nu) = \sum_{i,j} \nu(i, j) \log P_\theta(i, j). \tag{3.17}$$

Proof. Combine (3.9–11) and (3.13) with the definition of ν_K. Note that $X \in V(K, L)$ if f $\nu_K \in \hat{V}(K, L)$. One direction is obvious, the other follows from a classical theorem on the existence of Eulerian circuits on Eulerian graphs (see Kasteleyn [16]). □

Equation (3.14) tells us that $S(K, L)$ is the expectation of an exponential functional in level-2 form on the set $\hat{V}(K, L)$ where the underlying process is an i.i.d. sequence of geometric random variables. That is, we have achieved our first goal of this section. This reduction was possible because the probability of a path of a Markov process can be directly expressed in terms of its empirical pair distribution, which is why $\hat{P}(\nu)$ appears in the exponent in (3.14).

Remember that in order to prove Proposition 3 we must show that, in the notation of (3.7),

$$\lim_{n \to \infty} \frac{1}{n} \log E_\theta([\theta n], n) = J^\rho(\theta)$$

$$J^\rho(\theta) = \text{r.h.s.}(1.14).$$

By (3.12) this amounts to proving that

$$\lim_{n \to \infty} \frac{1}{n} \log S([\theta n], n) = J^\rho(\theta) + \theta \log(1 - c) + \tfrac{1}{2}(1 - \theta) \log c. \tag{3.18}$$

To see how this should come about, observe that for $\varepsilon > 0$ small and n large

$$\hat{V}([\theta n], n) = M_{\frac{[\theta n]}{n-1}} \subset \bigcup_{|\delta| \leq \varepsilon} M_{\theta + \delta}$$

with M_θ the set appearing in the variational formula (see 0.9). Now, if $\hat{V}([\theta n], n)$ were a *fixed* set and equal to M_θ and if M_θ were *closed* in the weak topology, then (3.18) would have immediately followed by applying Varadhan's Theorem ([20], Theorem 3.4) to (3.14). Indeed, both $\hat{F}$ and $\hat{P}$ are bounded and continuous on M_θ and the laws of v_K form a large deviation family with rate function

$$\hat{I}_c(v)=\sum_{i,j} v(i,j)\log\left(\frac{v(i,j)}{\bar{v}(i)(1-c)\,c^{j-1}}\right) \tag{3.19}$$

(cf. Ellis [9], p. 19). Therefore we would get

$$\lim_{n\to\infty}\frac{1}{n}\log S([\theta n],n)=\theta\sup_{v\in M_\theta}[\hat{F}(v)+\hat{P}(v)-\hat{I}_c(v)],$$

where the factor θ is picked up because $K=[\theta n]$. From there (3.18) would easily follow after substitution of (3.16) and (3.17) and after absorbing $\hat{P}(v)$ into $\hat{I}_c(v)$. (Note that the terms containing c can be computed because the two properties in the definition of M_θ imply that $\sum_{i,j}(j-1)\,v(i,j)=(1-\theta)/2\theta$). Incidentally, the form of $\hat{I}_c(v)$ in (3.19) can be deduced via a classical estimate on the number of Eulerian circuits on Eulerian graphs (see Kasteleyn [16]).

But unfortunately, $\hat{V}([\theta n],n)$ is *not fixed* and M_θ is *not closed*, and this is a serious problem. For instance, if we try to remove the first obstacle by replacing $\hat{V}([\theta n],n)$ in (3.14) by the slab

$$M_\theta^\varepsilon=\bigcup_{|\delta|\leqq\varepsilon} M_{\theta+\delta}$$

(assuming we could show that in the limit as $\varepsilon\to 0$ this has no effect on the growth rate of $S([\theta n],n)$), then we run into trouble because $\hat{F}$ and $\hat{P}$ are no longer continuous on M_θ^ε (and M_θ^ε is still not closed). If we try to repare this problem by passing to the natural stronger topology, namely the L_1-topology, then the level sets of $\hat{I}_c$ are no longer compact and so Varadhan's Theorem does not apply.

What we need is to collect more information on $\hat{F}$ and $\hat{P}$ in order to make the large deviation analysis possible. To do so we shall follow the traditional route of truncation of the X_k and combine this with a perturbation argument in θ. That is, we shall replace $\hat{V}([\theta n],n)$ in (3.14) by the set

$$M_\theta^{\varepsilon,R}=M_\theta^\varepsilon\cap\wp([1,R]^2)$$

and prove that in the limit as $\varepsilon\to 0$ and $R\to\infty$ this does not affect the growth rate of $S([\theta n],n)$. This will be carried out in Sect. 3c and 3d. In Sect. 3e we shall do the standard large deviation analysis on $M_\theta^{\varepsilon,R}$ and then prove that the resulting variational formula converges to the r.h.s. of (3.18) for $\varepsilon\to 0$ and $R\to\infty$.

3c Truncation

Our aim here is to show that for our purposes X can be restricted to a finite state space. For R positive integer let

$$\hat{V}^R(K,L)=\hat{V}(K,L)\cap\wp([1,R]^2)$$
$$S^R(K,L)=E(\exp\{K[\hat{F}(v_K)+\hat{P}(v_K)]\}\,\chi\{v_K\in\hat{V}^R(K,L)\}).$$

The main result of this section is the following lemma.

Lemma 6 *For every* $\theta \in (0,1)$

$$\lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n} |\log S([\theta n], n) - \log S^R([\theta n], n)| = 0. \tag{3.20}$$

Proof. The proof is in the spirit of that of Lemma 1 in Sect. 2a. The idea is to associate with each configuration X that somewhere exceeds the value R another configuration X' that everywhere stays below R and contributes about as much to the exponential. A configuration will be viewed as a collection of $[\theta n]$ piles of objects of sizes $1, X_1, \ldots, X_{[\theta n]-2}, 1$ and X' will be built out of X by moving objects around in small piles of a given size. Since all X have the same probability all objects may be moved around freely. The problem will be to control the effect on the exponential. In particular, it turns out that we have to be careful with those X where large piles occur next to each other. This will be handled by creating in X' many small piles in a row. The reason behind this is that the kernel P_θ in (3.17) belongs to branching with immigration, as described in Lemma 4, and therefore it has the property of favouring "flat" configurations.

Let $V(K, L)$ be the set of X's given by (3.9), and let

$$V^R(K, L) = V(K, L) \cap [1, R]^K.$$

It suffices to show that there exists $C(R)$ with $C(R) \to 0$ as $R \to \infty$ such that for all n

$$\begin{aligned} &E(\exp[F(X)+P(X)]\,\chi\{X \in V([\theta n], n) \setminus V^R([\theta n], n)\}) \\ &\leqq \exp(nC(R))\, E(\exp[F(X)+P(X)]\,\chi\{X \in V^R([\theta n], n)\}). \end{aligned} \tag{3.21}$$

Here F and P are given by (3.10) and (3.11). To get the lemma rewrite (3.21) in terms of the empirical pair distribution, as was done before to get (3.14) from (3.13).

Fix n and R. To prove (3.21) we need a map

$$T\colon\ V \setminus V^R \to V^R$$

which is defined as follows. Let s, t and u be positive integers such that

$$\begin{gathered} s, t, u \to \infty \\ u/s \to 0 \\ ut/s \to \infty \end{gathered}$$

and put $R = st$. Introduce the following subsets of $[0, [\theta n] - 1]$. Let

$$A^1 = \{k \colon X_k > st\}$$

and let A^2 be the smallest set containing A^1 such that

$$\begin{aligned} X_k &\geqq u(t-1) \quad \text{for } k \in A^2 \\ X_k &\leqq u(t-1) \quad \text{for } k \in \partial A^2 \end{aligned}$$

where ∂A^2 denotes the exterior boundary of A^2. Next, let

$$B^1 = \{k: X_k \leqq s(t-1), k \notin \partial A^2\}$$

and let B^2 be the subset of B^1 obtained by deleting all intervals in B^1 of length $<t$. Now T acts as follows:

> T removes from each $k \in A^2$ as many piles of size s until at most s objeccts remain. All the piles removed from A^2 are placed back on B^2, one pile on one site and filling B^2 in a row.

Note that the sets A^2 and B^2 need not necessarily be disjoint. Clearly, the image configuration is everywhere below st. Of course, for T to be well defined the set B^2 must be large enough to accomodate the piles coming from A^2. But this is so by the following observations. Since $\sum_k X_k = \frac{1}{2}(n + [\theta n] - 1) < n$,

$$|A^2| \leqq n/u(t-1)$$
$$|(B^1)^c| \leqq n/s(t-1) + |\partial A^2|.$$

Since at most $(t-1)|(B^1)^c|$ sites can lie inside the intervals in B^1 of length $<t$, and since $|\partial A^2| \leqq 2|A^2|$,

$$|B^2| > [\theta n] - n/s - 2n/u.$$

The total number of piles moved is at most n/s. So it suffices to have $|B^2| > n/s$, which holds by the above estimate as soon as $2/s + 2/u < [\theta n]/n \to \theta$.

We shall need four properties of T:

(3.22) $\quad X$ and TX have the same probability.

(3.23) $\quad F(X) - F(TX) \leqq n o(1)$ uniformly for $X \in V \setminus V^{st}$.

(3.24) $\quad P(X) - P(TX) \leqq n o(1)$ uniformly for $X \in V \setminus V^{st}$.

(3.25) $\quad |\{X \in V \setminus V^{st}: TX = X'\}| \leqq \exp(n o(1))$ uniformly for $X' \in V^{st}$.

Here $o(1)$ refers to s, t, $u \to \infty$. In the proof of these properties below we shall need the following observations. Let B^3 be the subset of B^2 where the piles are moved to. Then

$$\begin{aligned} (TX)_k &= X_k && \text{for } k \notin A^2 \cup B^3 \\ (TX)_k &\leqq s < X_k && \text{for } k \in A^2 \\ (TX)_k &= X_k + s && \text{for } k \in B^3, \end{aligned}$$

A^2 and B^3 are separated everywhere by at least one site, and $|B^3| < n/s$.

Proof of (3.23). First use (2.1) to write

$$F(X) - F(TX) = \sum_{k=1}^{[\theta n]} [g((TX)_{k-1} + (TX)_k - 1) - g(X_{k-1} + X_k - 1)].$$

If A^{2+} and B^{3+} denote the right closure of A^2 and B^3, then the above sum has two contributions namely $\sum_{k\in A^{2+}}$ and $\sum_{k\in B^{3+}}$. Since g is increasing the first sum is negative and can be trivially bounded above by zero. The second sum equals

$$\sum_{k\in B^{3+}} [g(X_{k-1}+X_k+\delta_k s-1)-g(X_{k-1}+X_k-1)]$$

with $\delta_k=1$ or 2 depending on whether k is in the interior of B^{3+} or not. Now use the concavity of g (2.1) to obtain

$$F(X)-F(TX)\leqq \sum_{k\in B^{3+}} g(\delta_k s)\leqq |B^{3+}|\, g(2s)\leqq 2ng(2s)/s.$$

Finish the proof with (2.1). □

Proof of (3.24). First substitute the special form (3.3) of the kernel to write

$$P(X)-P(TX)=\log \prod_{k=1}^{[\theta n]} \binom{X_{k-1}+X_k-2}{X_{k-1}-1}\binom{(TX)_{k-1}+(TX)_k-2}{(TX)_{k-1}-1}^{-1}.$$

Again there are two contributions namely $\prod_{k\in A^{2+}}$ and $\prod_{k\in B^{3+}}$. Now define

$$\Delta_k=X_k-(TX)_k.$$

The first product can be bounded above by dropping the second binomial coefficient, by using the inequality

$$\binom{a+b}{a}\leqq 2^{a+b}, \tag{3.26}$$

and by estimating

$$\begin{aligned}\sum_{k\in A^{2+}} [X_{k-1}+X_k-2]&\leqq 2\sum_{k\in A^2}(\Delta_k+s)+\sum_{k\in\partial A^2} ut\\ &\leqq 2\sum_{k\in A^2}\Delta_k+2ns/u(t-1)+2nu/s.\end{aligned}$$

Here we use that $|\partial A^2|\leqq 2|A^1|$ and $|A^1|\leqq n/st$. Hence the first product contributes to $P(X)-P(TX)$ at most

$$(2\log 2)\sum_{k\in A^2}\Delta_k+no(1).$$

Turning to the second product, by using the inequality

$$\binom{a}{b}\binom{c}{d}\leqq\binom{a+c}{b+d} \tag{3.27}$$

we get the upper bound

$$\prod_{k\in B^{3+}} \binom{-\Delta_{k-1}-\Delta_k}{-\Delta_{k-1}}^{-1}.$$

But $\Delta_k = -s$ for $k\in B^3$ and $\Delta_k = 0$ for $k\in\partial B^3$, and so this equals

$$\binom{2s}{s}^{-C}, \quad \text{with } C = |\{k: k-1, k\in B^3\}|.$$

Now note that B^3 consists of intervals of length at least t because the sites of B^2 are filled in a row (except possibly one interval where the filling of B^2 stops). Hence

$$|B^3| - C \leqq [t^{-1}|B^3|] + 1,$$

and hence the second product contributes to $P(X) - P(TX)$ at most

$$-\{[(1-t^{-1})|B^3|] - 1\} \log\binom{2s}{s}.$$

Finally, note that

$$\sum_{k\in A^2} \Delta_k = s|B^3| = \text{total number of objects moved}$$

and add the two contributions to arrive at

$$P(X) - P(TX) \leqq s|B^3| \left\{2\log 2 - (1-t^{-1})\, s^{-1} \log\binom{2s}{s}\right\} + n\,o(1).$$

The proof is now finished because the term between braces tends to zero and because $s|B^3| < n$. □

Proof of (3.25) The total number of piles moved is at most n/s. It follows that for every X' the number of X that are mapped onto X' cannot be more than

$$\{[\theta n]^{n/s}/(n/s)!\}^2$$

because there are at most $[\theta n]$ sites where a pile can be removed and where it can be placed back. Now use Stirling's formula. □

To complecte the proof of (3.21), first use (3.23) and (3.24) to get that the l.h.s. of (3.21) is bounded above by

$$\exp(n\,o(1))\, E(\exp[F(TX) + P(TX)]\, \chi\{X\in V([\theta n], n)\setminus V^R([\theta n], n)\}).$$

Then use (3.22) and (3.25) to bound this further by $\exp(n\,o(1))$ times the r.h.s. of (3.21). □

3d Perturbation: going to the slab

In this section we replace $\hat{V}^R([\theta n], n)$ by the slab $M_\theta^{\varepsilon, R}$. That is, instead of fixing $\sum_{k=0}^{[\theta n]-1} X_k$ at exactly the value $\frac{1}{2}(n+[\theta n]-1)$ we want to allow it to vary over a small slab of width εn. In order to achieve this we have to investigate how $S^R([\theta n], n)$ behaves under small perturbations of its second argument. The main result of this section is the following lemma. This will be used later to prove Lemma 8 below, which is the result that we are really after.

Lemma 7 *For every $\theta \in (0,1)$ there exists $C(\theta)$ such that for every $\varepsilon(\cdot)$: $\mathbb{N} \to \mathbb{Z}$ satisfying $|\varepsilon(n)| \leqq \varepsilon n$ for ε sufficiently small*

$$(3.28) \quad \frac{1}{n} |\log S^R([\theta n], n+\varepsilon(n)) - \log S^R([\theta n], n)| \leqq -C(\theta) \left| \frac{\varepsilon(n)}{n} \right| \log \left| \frac{\varepsilon(n)}{n} \right|.$$

Proof. Whether or not our random variables X_k are truncated plays no role at this point. We shall give the proof for the non-truncated X_k and explain later how to incorporate the truncation.

Return to (3.12) and (3.13). Define

$$(3.29) \quad \begin{aligned} U(K,L) &= E(\exp[F(X)+P(X)] \mid X \in V(K,L)) \\ &= S(K,L) |V(K,L)|^{-1} (1-c)^{-K+2-(L-K-1)/2}. \end{aligned}$$

We start with the observation that $U(K,L)$ is expectation with respect to the uniform distribution on $V(K,L)$ (recall that $V(K,L)$ fixes both length and sum of X and that the X_k are i.i.d. geometric). Therefore we introduce the Markov chain on N^K with transition kernel

$$Q(x,y) = x(i) \Big/ \sum_{j=1}^{K-2} x(j) \quad \text{for } y = x+\delta_i \quad \text{with } 1 \leqq i \leqq K-2, \quad \text{zero otherwise,}$$

and we observe that if this chain is started in the state $1 = (1, \ldots, 1)$, then its n-step distribution is the uniform distribution on $V(K, K+2n+1)$. Hence we can write

$$U(K,L) = \sum_X Q^{\frac{1}{2}(L-K-1)}(1,X) \exp[F(X)+P(X)].$$

This representation really is a trick to handle the combinatorics. Using that

$$|V(K,L)| = \binom{\frac{1}{2}(L+K-7)}{K-3}$$

and substituting the two cases

$$\begin{aligned} (K,L) &= ([\theta n], n+\varepsilon(n)) \\ (K,L) &= ([\theta n], n) \end{aligned}$$

we see from (3.29) that in order to prove the lemma we must show that $U(K, L)$ satisfies the estimate in (3.28). The other terms are exponential in n and contribute only $\varepsilon(n)/n$ to the estimate in (3.28), which is of lower order.

Now assume $\varepsilon(n)$ positive (the negative case is analogous) and write

$$\begin{aligned}(3.30)\qquad &\sum_X Q^{\frac{1}{2}(n-[\theta n]-1+\varepsilon(n))}(1, X) \exp[F(X)+P(X)] \\ &= \sum_{X'} Q^{\frac{1}{2}(n-[\theta n]-1)}(1, X') \exp[F(X')+P(X')] \\ &\quad \times \{\sum_X Q^{\frac{1}{2}(\varepsilon(n))}(X', X) \exp[F(X)-F(X')+P(X)-P(X')]\}.\end{aligned}$$

We shall show that for all X and X' on which the chain lives

$$(3.31)\qquad 2\varepsilon(n) f(1) \leqq F(X)-F(X') \leqq 2\varepsilon(n) \log M$$

$$\begin{aligned}(3.32)\qquad &0 \leqq P(X)-P(X') \\ &\leqq \varepsilon(n) \log[(1-\theta^2)] + \log\binom{n-[\theta n]-1+2\varepsilon(n)}{2\varepsilon(n)}.\end{aligned}$$

This gives uniform lower and upper bounds for the term between braces in (3.30), which then immediately become bounds on the ratio $U([\theta n], n+\varepsilon(n))/U([\theta n], n)$ after applying Stirling's formula to the binomial. This proves the lemma.

The key point is that by construction the chain only lives on X and X' for which

$$\begin{aligned}(3.33)\qquad &\sum_k X'_k = \tfrac{1}{2}(n+[\theta n]-1), \\ &\sum_k [X_k - X'_k] = \varepsilon(n), \\ &X_k \geqq X'_k \quad \text{for all } k.\end{aligned}$$

First use (2.1), by which f is convex, implying $f(i+1)-f(i) \geqq f(1)$, to get from (3.10)

$$F(X)-F(X') \geqq f(1) \sum_k [X_{k-1}+X_k-X'_{k-1}-X'_k] = 2 f(1)\, \varepsilon(n),$$

explaining the lower bound in (3.31). The upper bound follows as above from $f(i+1)-f(i) \leqq \log M$. To see (3.32) first use (3.3) and (3.11) to write

$$\begin{aligned}P(X)-P(X') &= \varepsilon(n) \log[\tfrac{1}{4}(1-\theta^2)] \\ &\quad + \log \prod_k \binom{X_{k-1}+X_k-2}{X_{k-1}-1}\binom{X'_{k-1}+X'_k-2}{X'_{k-1}-1}^{-1}.\end{aligned}$$

The quotient of binomials is easily seen to be bounded above by the product

$$\binom{X_{k-1}+X_k-2}{\Delta X_{k-1}+\Delta X_k}\binom{\Delta X_{k-1}+\Delta X_k}{\Delta X_{k-1}},$$

with $\Delta X_k = X_k - X_k'$. The product over k of the first factor can be bounded above by iterating (3.27) and using (3.33). This gives the binomial in the r.h.s. of (3.32). The product over k of the second factor can be bounded above by applying (3.26) termwise. This gives a term $2^{2\varepsilon(n)}$. After collecting the various contributions we get the upper bound in (3.32). The lower bound is trivial, because of (3.27). This completes the proof.

It remains to see how the proof can be modified when we deal with truncated X_k. But this is easy: simply restrict the sum over X and X' to $[1,R]^K$ in the definition of $U(K,L)$ and in (3.30). The rest is the same. Incidentally, observe that the truncating map T that was used there preserves the sum of the X_k. So if we increase this sum by $\varepsilon(n)$ then nothing changes in the argument, and for the estimates in Sect. 3.c all that was needed is that the sum is of order n anyway. In other words, we are free to interchange the order of truncation and perturbation. □

We can now formulate the lemma that is the equivalent of Lemma 6. Let

(3.34)
$$M_\theta^{\varepsilon,R} = \{\nu \in \wp([1,R]^2) \colon \sum_j \nu(i,j) = \sum_j \nu(j,i), \sum_{i,j} (i+j-1)\,\nu(i,j) \in [\theta^{-1}(1-\varepsilon), \theta^{-1}(1+\varepsilon)]\}$$
$$S^{\varepsilon,R}(K) = E(\exp\{K[\hat F(\nu_K) + \hat P(\nu_K)]\}\,\chi\{\nu_K \in M_\theta^{\varepsilon,R}\}).$$

Lemma 8 *For every* $\theta \in (0,1)$

(3.35)
$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} |\log S^R([\theta n], n) - \log S^{\varepsilon,R}([\theta n])| = 0.$$

Proof. Simply note that for $\varepsilon > 0$ small and all n sufficiently large

$$\{\nu_{[\theta n]} \in \hat V^R([\theta n], n)\} \subset \{\nu_{[\theta n]} \in M_\theta^{\varepsilon,R}\} \subset \bigcup_{|i| \leqq 2\varepsilon n} \{\nu_{[\theta n]} \in \hat V^R([\theta n], n+i)\}.$$

Hence

$$S^R([\theta n], n) \leqq S^{\varepsilon,R}([\theta n]) \leqq \sum_{|i| \leqq 2\varepsilon n} S^R([\theta n], n+i).$$

Now apply Lemma 7. □

3e Large deviation analysis

The purpose of this section is to prove (3.18) and thereby finish the proof of Proposition 3. Having prepared for this in Sect. 3.c and 3.d, we start by giving the related result for the truncated random variables on the slab.

Lemma 9 *For every* $\theta \in (0,1)$, *and with* $\hat I_c(\nu)$ *as defined in* (3.19),

(3.36)
$$\lim_{n\to\infty} \frac{1}{n} \log S^{\varepsilon,R}([\theta n]) = \theta \sup_{\nu \in M_\theta^{\varepsilon,R}} [\hat F(\nu) + \hat P(\nu) - \hat I_c(\nu)].$$

Proof. I.i.d. geometric random variables satisfy the level-2 large deviation principle for pairs with rate function $\hat{I}_c(\nu)$ on the set $\wp(\mathbb{N}^2)$. Now, $I_c(\nu)$ is continuous on $\wp([1,R]^2)$ and $M_\theta^{\varepsilon,R}$ is a closed subset of $\wp([1,R]^2)$ equipped with the weak topology of measures. The latter facts imply via a standard argument that we also have a large deviation principle on $M_\theta^{\varepsilon,R}$ with the same rate function. Since both $\hat{F}$ and $\hat{P}$ are bounded and continuous on $M_\theta^{\varepsilon,R}$, (3.36) immediately follows by applying Varadhan's Theorem to (3.34), ([20], Theorem 3.4.). Recall that the factor θ is picked up because $K=[\theta n]$. □

The last piece of the puzzle is Lemma 10 below. Lemmas 6 and 8–10 imply (3.18) and therefore complete the proof of Proposition 3. □

Lemma 10 *For every* $\theta\in(0,1)$

$$\sup_{R<\infty}\inf_{\varepsilon>0}\sup_{\nu\in M_\theta^{\varepsilon,R}}[\hat{F}(\nu)+\hat{P}(\nu)-\hat{I}_c(\nu)] = \sup_{\nu\in M_\theta}[\hat{F}(\nu)+\hat{P}(\nu)-\hat{I}_c(\nu)]. \tag{3.37}$$

Proof. In this proof we abbreviate

$$K(\nu)=\hat{F}(\nu)+\hat{P}(\nu)-\hat{I}_c(\nu).$$

The first step is to show that for all R

$$\inf_{\varepsilon>0}\sup_{\nu\in M_\theta^{\varepsilon,R}}K(\nu)=\sup_{\nu\in M_\theta^R}K(\nu), \tag{3.38}$$

where of course $M_\theta^R=M_\theta\cap\wp([1,R]^2)$. This is a standard compactness argument. Indeed, for every $\varepsilon>0$ pick a measure $\nu_\varepsilon\in M_\theta^{\varepsilon,R}$ with

$$K(\nu_\varepsilon)\geqq -\varepsilon+\sup_{\nu\in M_\theta^{\varepsilon,R}}K(\nu).$$

Let (ν_{ε_k}) be any subsequence of (ν_ε) along which $K(\nu_\varepsilon)$ converges to its limsup as $\varepsilon\to 0$, and let ν be any weak limit point of (ν_{ε_k}). Then, since $K(\cdot)$ is continuous on $M_\theta^{\varepsilon,R}$ and $\nu\in M_\theta^R$, it follows that

$$\limsup_{k\to\infty}K(\nu_{\varepsilon_k})=K(\nu)\leqq\sup_{\nu\in M_\theta^R}K(\nu).$$

This implies (3.38) because $M_\theta^R\subset M_\theta^{\varepsilon,R}$ for all ε.

The second step is to show that

$$\sup_{R<\infty}\sup_{\nu\in M_\theta^R}K(\nu)=\sup_{\nu\in M_\theta}K(\nu). \tag{3.39}$$

Now, for every $\nu \in M_\theta$ there exists a sequence (ν_R) with $\nu_R \in M_\theta^R$ and $\nu_R \to \nu$ weakly as $R \to \infty$. Below we prove that $K(\cdot)$ is continuous on M_θ and so it follows that

$$\lim_{R\to\infty} K(\nu_R) = K(\nu).$$

This implies (3.39) because $M_\theta \supset M_\theta^R$ for all R.

To get the continuity of $K(\cdot)$ on M_θ, pick ν, $\nu_k \in M_\theta$ such that $\nu_k \to \nu$ weakly (as $k \to \infty$). Note that the restriction

$$\sum_{i,j} (i+j-1)\, \nu_k(i,j) = \theta^{-1} \qquad \text{for all } k$$

implies that ν_k sums $(i,j) \to i+j-1$ uniformly. Since both $f(i+j-1)$ and $\log P_\theta(i,j)$ in (3.16) and (3.17) are bounded by a constant times $i+j$, as is easily seen by using (2.1), (3.3) and (3.26), it follows that $\hat{F}$ and $\hat{P}$ are continuous. We finish the proof by showing that also $\hat{I}_c$ is continuous.

Recall (3.19). Consider first the entropy of ν_k

$$-\sum_{i,j} \nu_k(i,j) \log \nu_k(i,j). \tag{3.40}$$

Pick a $\gamma > 0$ and split the sum into two parts, running over the index sets J_k and J_k^c with

$$J_k = \{(i,j):\ \nu_k(i,j) \leqq (i+j)^{-\gamma}\}.$$

Pick a $\delta \in (0,1)$. On J_k

$$\begin{aligned} 0 &\leqq -\nu_k(i,j) \log \nu_k(i,j) \leqq C_\delta [\nu_k(i,j)]^{1-\delta} \\ &\leqq C_\delta (i+j)^{-(1-\delta)\gamma} \qquad \text{for some } C_\delta > 0. \end{aligned}$$

The second inequality holds because $-x^\delta \log x \to 0$ as $x \to 0$. On J_k^c

$$0 \leqq -\nu_k(i,j) \log \nu_k(i,j) \leqq \gamma\, \nu_k(i,j) \log(i+j).$$

It follows that $\nu_k(i,j) \log \nu_k(i,j)$ is uniformly summable as soon as $\gamma(1-\delta) > 2$ and hence (3.40) converges to the entropy of ν. Of course, the same applies for the entropy of $\bar{\nu}_k$ and $\bar{\nu}$. □

Having completed the proof of Proposition 3 we are now in a position to settle two old debts, namely (1.11) and (2.12). That is, we have to check that for every $\theta \in [0,1]$

$$\begin{gathered} E_{\theta_n}([\theta_n n], n) \qquad (\theta_n \to \theta) \\ E_\theta([\theta n] + 2\delta(n), n + 2\delta(n)) \qquad (\delta(n) = o(n)) \end{gathered}$$

have the same growth rate as $E_\theta([\theta n], n)$. But this is now straightforward. Indeed, from the analysis in Sect. 3e it is clear that the perturbations in K and L (in the notation of (3.7)), are negligible because they are $o(n)$. So all that is left to do is to remove the perturbation in the index θ. From (3.3) we get that

$$P_{\theta_n}(i,j)/P_\theta(i,j)=\left(\frac{1+\theta_n}{1+\theta}\right)^i\left(\frac{1-\theta_n}{1-\theta}\right)^{j-1}.$$

Via (3.8–11) one easily checks that

$$E_{\theta_n}([\theta_n n], n)/E_\theta([\theta n], n)=\exp(o(n))$$

if $\theta_n \to \theta$. $\square$

4 Proof of Proposition 4

This section is devoted to the proof of (1.15) and parallels Sect. 2. As there, the aim is to show that on an exponential scale the expectations in (1.15) are determined by those paths that stay between 0 and $[\theta n]$ until time n, cross $[\theta n]$ at time n, and then stay to the right of $[\theta n]$ forever. Again, the main idea will be to construct a map between the class of such paths and its complement in order to compare the respective contributions.

We retain the notation of Sect. 2. The only difference is that $-G_n(S)$ defined in (2.5) is replaced by

(4.1)
$$F_n(\omega, S)=\sum_{x=-\infty}^{\infty} l_n(x, S)\log(b_x)$$

where

$$\omega=\{b_x\}_{x=-\infty}^{\infty}$$

and that the function $\delta\colon \mathbb{N}\to\mathbb{N}$ in (2.2) is chosen such that

(4.2)
$$\begin{aligned}&\delta(n)\to\infty\\&\delta(n)=o(\log n).\end{aligned}$$

Our starting point is the equivalent of (2.4)

(4.3)
$$\text{l.h.s.}(1.15)=E_\theta(\exp[F_n(\omega, S)]\,\chi\{S\in A_n^1\}).$$

We emphasize that throughout this section ω is a *fixed* realization of the random environment and therefore $F_n(\omega, S)$ is a random variable only in S.

For simplicity we assume throughout Sects. 4 and 5 that $\operatorname{supp}\beta$ consists of *finitely* many points. At the end of Sect. 5 we shall argue that it is easy to eliminate this restriction.

Lemma 11 $E_\theta(\exp[F_n(\omega, S)]\,\chi\{S\in A_n^1\})\cong E_\theta(\exp[F_n(\omega, S)]\,\chi\{S\in A_n^2\})$ ω-a.s.

Proof. It suffices to show that

$$E_\theta(\exp[F_n(\omega, S)]\,\chi\{S\in A_n^1\backslash A_n^2\})\lesssim E_\theta(\exp[F_n(\omega, S)]\,\chi\{S\in A_n^2\})\quad \omega\text{-a.s.}$$

As in the proof of Lemma 1 we shall achieve this by comparing contributions of paths in $A_n^1\backslash A_n^2$ and A_n^2 via a map. Define

(i) a *small left crossing* over $[x-\delta(n), x]$ is a path from x to $x-\delta(n)$ staying inside $(x-\delta(n), x+\delta(n))$;
(ii) a *small right crossing* over $[x, x+\delta(n)]$ is a path from x to $x+\delta(n)$ staying inside $(x-\delta(n), x+\delta(n))$.
The idea in Sect. 2 was the following. A path in $A_n^1 \setminus A_n^2$ has excursions to the left of 0 and to the right of $[\theta n]$ that take it onto or beyond the boundaries of J_n. These we called large excursions and we used a map to bring the large excursions inside J_n by reflection around 0 and $[\theta n]$, thus getting an image path in A_n^2. Now, however, ω appears in $F_n(\omega, S)$ and since ω is spatially varying we shall have to be more subtle in bringing the large excursion inside J_n. Here is how we proceed.

Any large excursion can be decomposed into small crossings over the intervals

$$\begin{aligned} K_n(k) &= [k\delta(n), (k+1)\delta(n)] && (k<0) \\ K_n(k) &= [[\theta n]+k\delta(n), [\theta n]+(k+1)\delta(n)] && (k \geqq 0) \end{aligned}$$

by cutting the path at the stopping times

$$\begin{aligned} \tau_0 &= 0 \\ \tau_{i+1} &= \inf\{j > \tau_i : |S_j - S_{\tau_i}| = \delta(n)\} \qquad (i \geqq 0). \end{aligned}$$

Number the small crossings in order of appearance. For each k group together *consecutive* small crossings over $K_n(k)$. These will be called *strings* over $K_n(k)$. Each string over $K_n(k)$ is a piece of the path *separate in time* staying inside the interval of length $3\delta(n)-1$ around $K_n(k)$ defined by

$$\bar{K}_n(k) = \{x : |x - K_n(k)| < \delta(n)\}.$$

Note that by continuity of the path the numbers of small left and right crossings over $K_n(k)$ are equal so that the strings over $K_n(k)$ can be combined to form a *loop*.

Call an element of $(\operatorname{supp}\beta)^{3\delta(n)-1}$ a *type*. Each interval $(x, x+3\delta(n))$ carries a type for the given ω. Denote the type of $\bar{K}_n(k)$ by $t_n(k) = \{b_x\}_{x \in \bar{K}_n(k)}$. The following fact will be important:

> ω-a.s. there exists $n_0 = n_0(\omega)$ such that for every $n \geqq n_0$ there is a collection of intervals $L_n^t = [x^t, x^t + \delta(n)]$ indexed by $t \in (\operatorname{supp}\beta)^{3\delta(n)-1}$ such that $\bar{L}_n^t = \{x : |x - L_n^t| < \delta(n)\}$ carries type t and $\bar{L}_n^t$ are disjoint and are contained in $\operatorname{int}(J_n)$.

In other words, all types are represented in $\operatorname{int}(J_n)$ over disjoint intervals of length $3\delta(n)-1$. It is straightforward to deduce this fact from the i.i.d. property of ω using the assumption $\delta(n) = o(\log n)$ and $\operatorname{supp}\beta$ finite.

Now we are ready to define for $n \geqq n_0$ a map

$$T_n : A_n^1 \setminus A_n^2 \to A_n^2$$

acting on the path S between time 0 *and n*:

> For each k all strings over $K_n(k)$ are cut out of the path and made into a loop in order of appearance. This loop is moved to $L_n^{t_n(k)}$ and fitted into the path immediately after the last hitting time of the right (left) boundary of $L_n^{t_n(k)}$ if $k<0$ ($k \geqq 0$). This procedure is done successively for k running from $-\infty$ to ∞.

We shall show that T_n has the following three properties:

(4.4) T_n preserves probability: any n-step path and its image under T_n have the same probability;

(4.5) $\qquad F_n(\omega, T_n(S)) = F_n(\omega, S) \qquad$ for all $S \in A_n^1 \backslash A_n^2$,

(4.6) $\qquad |\{S \in A_n^1 \backslash A_n^2 : T_n(S) = S'\}| = \exp(o(n)) \qquad$ uniformly for $S' \in A_n^2$.

Both (4.4) and (4.5) are obvious from the construction. In particular, note that $F_n(\omega, S)$ only depends on the total local times in the level sets of ω and that these are not changed by the map T_n because strings are moved only between intervals of length $3\delta(n)-1$ carrying the same type. We now come to the nontrivial part (4.6).

Proof of (4.6). We shall argue that

$$|\{S \in A_n^1 \backslash A_n^2 : T_n(S) = S'\}| \leqq (n/2\delta(n))^2 \, 2^{n/\delta(n)}.$$

Pick an $S' \in A_n^2$. We are going to count in how many ways we can construct an $S \in A_n^1 \backslash A_n^2$ having S' as its image. Decompose S' into small crossings and single out the ones over the collection (L_n^t). The latter are the only candidates for being images of small crossing in S moved by T_n. By continuity of the path, S cannot reach intervals $K_n(k)$ with $|k| > n/2\delta(n)$ between time 0 and n. Therefore there are no more than $(n/2\delta(n))^2$ ways to choose the left most $K_n(k^-)$ and the right most $K_n(k^+)$ over which S has small crossings (and by continuity S must have small crossings over every $K_n(k)$ in between). Now let k run through $[k^-, k^+]$ and for each k take away from S' an equal (positive) number S of small left and right crossings over $L_n^{t_n(k)}$. Group them together into a number of strings over $L_n^{t_n(k)}$ in order of appearance and move them back to $K_n(k)$. The number of ways in which this can be done is at most the number of ways in which the small crossings over (L_n^t) in S' can be divided into (nonempty) groups. Since there are no more than $n/\delta(n)$ crossings altogether this gives the upper bound $2^{n/\delta(n)}$. Once the strings are moved back there are two possibilities: either they can be fitted together to get a continuous path S or they cannot. In the first case S is uniquely determined because strings are ordered and are separate in time over each $K_n(k)$: at the end of each string it is uniquely determined what string the path S must follow next. In the second case we have overestimated. This completes the proof of (4.6). $\square$

We can now finish the proof of Lemma 11. Applying first (4.5) and then (4.4) and (4.6), we obtain

$$\begin{aligned} E_\theta(\exp[F_n(\omega, S)]\,\chi\{S \in A_n^1 \backslash A_n^2\}) &= E_\theta(\exp[F_n(\omega, T_n(S))]\,\chi\{S \in A_n^1 \backslash A_n^2\}) \\ &\gtrsim E_\theta(\exp[F_n(\omega, S')]\,\chi\{S' \in A_n^2\}), \end{aligned}$$

which is what we set out for. $\square$

The rest of this section is a copy of Sect. 2. The following two lemmas are the analogues of Lemmas 2 and 3. Define, to replace $G_n^*(S)$ in (2.9),

$$F_n^*(\omega, S) = \sum_{x \in J_n} l(x, S) \log(b_x). \tag{4.7}$$

Lemma 12 $E_\theta(\exp[F_n(\omega, S)]\, \chi\{S \in A_n^2\}) \cong E_\theta(\exp[F_n(\omega, S)]\, \chi\{S \in A_n^3\})$ ω-a.s.

Lemma 13 $E_\theta(\exp[F_n(\omega, S)]\, \chi\{S \in A_n^3\}) \cong E_\theta(\exp[F_n^*(\omega, S)]\, \chi\{S \in A_n^3\})$ ω-a.s.

Proof. The proof is literally the same as in Sect. 2. □

With these preparations we are ready to prove (1.15). Lemmas 11–13 combine with (4.3) to give the upper bound along the same lines as in (2.10–12). The lower bound as in (2.13) is again trivial. □

5 Proof of Proposition 5

This section is devoted to the proof of (1.16) and is in structure similar to Sect. 3, where we proved (1.14). We assume the reader is familiar with the basic ideas there. Like (1.14), the expectation in (1.16) involves an exponential functional of the total local time process $\{l(x)\}$ in the presence of an indicator on the sum. The difference is that $f(l(x))$ in (1.14) is replaced by $f_x(l(x)) = l(x) \log(b_x)$ in (1.16), which now also depends on the random environment ω through the mean offspring b_x at site x. This will bring about several modifications in the proof, some of which are technical but some of which require new ideas. In particular, we shall now have to deal with the empirical pair distribution of the Markov process and the random environment combined, i.e. $\{m(x), b_x\}$ (see (3.1–4)), and we shall have to do a large deviation analysis on level -3 since ω is fixed. For this new situation we derive the large deviation principle and compute the entropy function and obtain finally the variational formula for the exponential growth rate of the exponential functional in (1.16).

Section 5.a replaces the preparatory work in Sect. 3b–d while Sect. 5.b contains the large deviation analysis analogous to Sect. 3e. Many of the tools of Sect. 3 will reappear and this will allow for fairly short proofs.

5a Preparations

In order to analyze the expectation in the l.h.s. of (1.16) we retain the definitions (3.7–13) with the understanding that $f(l(x))$ is everywhere replaced by $l(x) \log(b_x)$. This means that throughout this section we use the same symbols as in Sect. 3 even though some of them will now acquire a different meaning, as we shall point out along the way. Since the structure of the arguments is very similar to Sect. 3, we choose this misuse of notation rather than the option of inventing many new symbols. The basic rule of translation is to replace the number X_k by the vector (X_k, b_k). The key equations are (3.7–13) and we assume the reader understands how to read these equations in the new context.

To be able to now rewrite (3.13) in level-2 form we introduce

$$\nu_K = K^{-1} \sum_{k=1}^{K} \delta_{((X_{k-1}, b_{k-1}), (X_k, b_k))}.$$

This is the empirical pair distribution of the *vector-valued* process $\{(X_k, b_k)\}$, where X_k are the i.i.d. geometric random variables introduced in Sect. 3.b and b_k are the fixed mean offsprings, both indexed by site k. Recall that we are using periodic boundary conditions $(X_K, b_K) = (X_0, b_0)$ and that $X_0 = X_{K-1} = 1$. Also recall that β is assumed to have finite support.

Our starting point is the following analogue of Lemma 5. Let $S_\omega(K, L)$ denote the new version of (3.13). We add the index ω to remind the reader that the medium is fixed and that the symbol should be read in the new context. E_ω denotes expectation over $\{X_k\}$ for fixed ω. We abbreviate $i = (i_1, i_2)$ and $j = (j_1, j_2)$ with $i_1, j_1 \in \mathbb{N}$ and $i_2, j_2 \in \operatorname{supp} \beta$. Recall also that $K = [\theta n]$ and $L = n$.

Lemma 14 *For every* $\theta \in (0, 1)$

$$S_\omega(K, L) = E_\omega(\exp\{K[\hat{F}(\nu_K) + \hat{P}(\nu_K)]\} \chi\{\nu_K \in \hat{V}(K, L)\}) \tag{5.1}$$

where

$$\begin{aligned}\hat{V}(K, L) = \{\nu \in \wp((\mathbb{N} \otimes \operatorname{supp} \beta)^2):&\\ \sum_j \nu(i, j) = \sum_j \nu(j, i) \quad &\textit{for all } i \in \mathbb{N},\\ \sum_{i,j} (i_1 + j_1 - 1)\, \nu(i, j) = L - 1)/K\}&\end{aligned} \tag{5.2}$$

$$\hat{F}(\nu) = \sum_{i,j} \nu(i, j)(i_1 + j_1 - 1) \log j_2 \tag{5.3}$$

$$\hat{P}(\nu) = \sum_{i,j} \nu(i, j) \log P_\theta(i_1, j_1). \tag{5.4}$$

Proof. By introducing the vectors i and j the proof becomes literally the same as that of Lemma 5. □

The rest of this section consists of four steps: (1) truncation, (2) perturbation and passing to the slab, (3) showing that the growth rate of $S_\omega([\theta n, n])$ is a.s. independent of ω, (4) integration over ω that are typical for the medium. The last two steps are where we do something new compared to Sect. 3. The last step is very important because it gives us an expectation over the *double process* $\{m(x), b_x\}$, for which we can deduce the level-3 large deviation principle by standard arguments.

Step 1. First we truncate, i.e. the state space $\mathbb{N}$ of X_k is replaced by the finite set $[1, R]$. The necessity of this approach was explained in the last two paragraphs of Sect. 3b.

Lemma 15 *Lemma 6 continues to hold when* $[1, R]$ *is replaced by* $[1, R] \otimes \operatorname{supp} \beta$ *and* $\hat{V}(K, L)$, $\hat{F}(\nu)$, $\hat{P}(\nu)$ *by* (5.2–4) *and* $S(K, L)$, $S^R(K, L)$ *by* $S_\omega(K, L)$, $S^R_\omega(K, L)$.

Proof. What we must do is prove (3.21) in the new context for all $n \geq n_0 = n_0(R, \omega)$. The idea in the proof of Lemma 6 was to view $X =$

$(1, X_{1'}, \ldots, X_{[\theta n]-2}, 1)$ as a collection of $[\theta n]$ piles, to truncate large piles by splitting them into small piles and then moving the small piles to a string of consecutive sites. We shall do the same here, except that now there is the additional complication that we shall want to move small piles *only within the level sets of* ω in order to control the effect on the exponential. This means that we shall want to change the definition of the map T introduced below (3.21).

Here is how the proof of Lemma 6 will be modified. Retain the definition of the sets A^1, A^2, B^1 and B^2. Let

$$C_b^1 = \{k \in B^2 : b_k = b\} \qquad (b \in \operatorname{supp} \beta)$$

be the level sets of ω in B^2 and let C_b^2 be the subset of C_b^1 obtained by deleting all intervals in C_b^1 of length $< \kappa \log t$. The constant κ will be chosen later. Now define the map T as follows:

> T removes from each $k \in A^2$ as many piles of size s until $(TX)_k \leqq s$.
> For each $b \in \operatorname{supp} \beta$ all the piles removed from $\{k \in A^2 : b_k = b\}$
> are placed back on C_b^2, one pile on one site and filling C_b^2 in a row.

If

$$|C_b^2| > n/s \qquad \text{for all } b \in \operatorname{supp} \beta, \tag{5.5}$$

then T is well defined because no more than n/s piles are moved. Therefore we have to see that for all n large and ω-a.s. this condition is satisfied.

The set B^2 consists of intervals of length $\geqq t$ and therefore contains at least $|B^2|/2t$ disjoint intervals of length t. Let

> $\alpha(\kappa, t) =$ probability that a given interval of length t contains
> intervals of length $\geqq \kappa \log t$ in each of the level sets of ω.

Because of the i.i.d. property of ω the strong law implies that ω-a.s. there exists $n_0 = n_0(\omega, s, t, u)$ such that for $n \geqq n_0$ the number of intervals of length $\geqq \kappa \log t$ in each of the level sets of ω in B^2 exceeds $\alpha(\kappa, t)|B^2|/4t$. Since $|B^2| > [\theta n] - n/s - 2n/u$, as shown in the proof of Lemma 6, it follows that

$$|C_b^2| \geqq \frac{\alpha(\kappa, t)\, \kappa \log t}{4t} ([\theta n] - n/s - 2n/u) \qquad \text{for all } b \in \operatorname{supp} \beta.$$

Now, the i.i.d. property of ω implies that

$$\lim_{t \to \infty} \alpha(\kappa, t) = 1 \qquad \text{for } \kappa \text{ small enough.}$$

Therefore to get (5.5) we must require $n \geqq n_0$ and

$$\begin{gathered} s, t, u \to \infty \\ (s \log t)/t \to \infty. \end{gathered} \tag{5.6}$$

For the rest of the proof of Lemma 6 to carry over we want that properties (3.22–25) continue to hold. This requires that

$$\begin{aligned} u/s &\to 0 \\ (u \log t)/s &\to \infty. \end{aligned} \tag{5.7}$$

In fact, (3.23) is now trivial because piles are only moved within level sets of ω, so that $F(X)=F(TX)$, while (3.24) is the same as before because $P(X)$ does not depend on ω, except that now t is replaced by $\kappa \log t$ below (3.26). This completes the proof of the analogue of Lemma 6: to match (5.6) and (5.7) choose for example $s=t$, $u=t/\log\log t$. □

Step 2. Next we perturb and go to the slab

$$\begin{aligned} M_\theta^{\varepsilon,R} = \{ v \in \wp(([1,R] \otimes \operatorname{supp} \beta)^2)\colon & \\ \sum_j v(i,j) = \sum_j v(j,i) \quad & \text{for all } i \in \mathbb{N}, \\ \sum_{i,j} (i_1 + j_1 - 1)\, v(i,j) \in [\theta^{-1}(1-\varepsilon), & \theta^{-1}(1+\varepsilon)] \}, \end{aligned}$$

i.e. $\hat{V}([\theta n], n)$ is replaced by $M_\theta^{\varepsilon,R}$ in (5.1) (recall (3.34)).

All of Sect. 3.d carries over except for (3.31), where now only the lower bound changes and becomes $2\varepsilon(n) \log m$ with m the minimal value in $\operatorname{supp} \beta$. Hence we have

Lemma 16 *Lemmas 7 and 8 continue to hold when* $[1,R]$ *is replaced by* $[1,R] \otimes \operatorname{supp} \beta$ *and* $\hat{V}(K,L)$, $\hat{F}(v)$, $\hat{P}(v)$ *by* (5.2–4) *and* $S^R(K,L)$, $S^{\varepsilon,R}(K)$ *by* $S_\omega^R(K,L)$, $S_\omega^{\varepsilon,R}(K)$.

Step 3. The third step is a very crucial one: we show that the limiting growth rate of $S_\omega^{\varepsilon,R}([\theta n])$ is a.s. independent of ω.

Lemma 17 *For every* $\theta \in (0,1)$ *and for almost all* ω, ω'

$$\limsup_{n\to\infty} \frac{1}{n} |\log S_\omega^{\varepsilon,R}([\theta n]) - \log S_{\omega'}^{\varepsilon,R}([\theta n])| = 0. \tag{5.8}$$

Proof. Our task is to compare the effect of conditioning on ω or ω'. Both are typical realizations of the random environment and therefore have the same statistics. Since $X=\{X_k\}$ is i.i.d. the idea will be to permute the X_k in (X,ω) so as to imitate (X,ω') asymptotically. Here is how we proceed.

Fix a positive integer N and divide ω and ω' into blocks of length N. The possible values of $(b_{mN+k})_{k=1}^N$ and $(b'_{mN+k})_{k=1}^N$ define the *type* of the N-block with number $m=0,1,2,\ldots$. Now define a map T which acts on X as follows:

> Look at block 0 in ω and read off its type. Look for the first block in ω' of the same type. Say this is block m_0. Then replace $(X_k)_{k=1}^N$ by $(X_{m_0N+k})_{k=1}^N$. Now apply the same procedure to block 1 in ω by looking for the first block $m_1 \neq m_0$ in ω' of the same type and replacing $(X_{N+k})_{k=1}^N$ by $(X_{m_1N+k})_{k=1}^N$, etc.

(T is well defined ω-a.s.) Thus TX is obtained by shuffling N-blocks of X, where block i is replaced by block m_i with $\{m_i\}$ a permutation (depending on ω, ω', N). The following properties are obvious from the construction:

(5.9) T is one-to-one and preserves probability for almost all ω, ω'.

(5.10) Let ν_K and ν'_K be the empirical pair distributions of (TX, ω) and (X, ω') over the interval $[1, K]$. For K a multiple of N let

$$\tilde{\mu}_K^N(\omega) = \frac{N}{K} \sum_{m=0}^{K/N-1} \delta_{(b_{mN+k})_{k=1}^N}$$

be the empirical distribution of N-blocks in ω over $[1, K]$. For $\delta > 0$ let

$$B^{\delta, N} = \{\tilde{\mu} \in \wp((\operatorname{supp} \beta)^N) \colon \|\tilde{\mu} - \beta^N\| \leqq \delta\}$$

($\|\cdot\|$ denotes total variation). Then for all ω, ω' such that $\tilde{\mu}_K^N(\omega), \tilde{\mu}_K^N(\omega') \in B^{\delta, N}$

$$\|\nu_K - \nu'_K\| \leqq \frac{1}{N} + 2\delta$$
$$\Big|\sum_{i,j} (i_1 + j_1 - 1)(\nu_K(i,j) - \nu'_K(i,j))\Big| \leqq 4\delta R.$$

Observe that by the strong law for every N and δ there exists $n_0 = n_0(\omega, \omega', N, \delta)$ such that $\tilde{\mu}_K^N(\omega), \tilde{\mu}_K^N(\omega') \in B^{\delta, N}$ for $n \geqq n_0$ (recall that $K = [\theta n]$).

Now argue as follows. Observe that $S_\omega^{\varepsilon, R}([\theta n])$ is an expectation over X which by (5.9) is invariant under replacing X by TX. Then use (5.10) in combination with Lemma 16 (noting that the latter is a statement about perturbations on the restriction $\sum_{i,j} (i_1 + j_1 - 1)\, \nu_K(i,j) = (L-1)/K$) to conclude that for $n \geqq n_0$

$$\begin{aligned} &\frac{1}{n} |\log S_\omega^{\varepsilon, R}([\theta n]) - \log S_{\omega'}^{\varepsilon, R}([\theta n])| \\ &\quad \leqq \frac{[\theta n]}{n} \Big(\frac{1}{N} + 2\delta\Big) M - C(\theta) \Big(\frac{[\theta n]}{n} 4\delta R\Big) \log \Big(\frac{[\theta n]}{n} 4\delta R\Big), \end{aligned}$$

($C(\theta)$ is the constant in Lemma 7). Now take $n \to \infty$, $\delta \to 0$ and $N \to \infty$. □

Step 4. Finally we integrate ω over the event $\{\tilde{\mu}_K^N(\omega) \in B^{\delta, N}\}$ and we display $S_\omega^{\varepsilon, R}(K)$ as a function of

$$\mu_K^N = \frac{N}{K} \sum_{m=0}^{K/N-1} \delta_{(X_{mN+k}, b_{mN+k})_{k=1}^N},$$

the empirical distribution of N-blocks over $[1, K]$ of the double process $\{(X_k, b_k)\}$. Let ϑ be the cyclic shift on vector-valued coordinates and π^2 the projection on the first two of the N vector-valued coordinates. Let $\sim$ denote projection on the ω-coordinates (e.g. $\mu_K^N \to \tilde{\mu}_K^N(\omega)$).

Lemma 18 *For every $\theta \in (0,1)$ and for almost all ω*

(5.11)
$$\lim_{\substack{N\to\infty\\ \delta\to 0}} \limsup_{n\to\infty} \frac{1}{n} |\log S_\omega^{\varepsilon,R}([\theta n]) - \log S^{\delta,N,\varepsilon,R}([\theta n])| = 0$$

where

(5.12)
$$S^{\delta,N,\varepsilon,R}(K) = \int S_\omega^{\varepsilon,R}(K)\, \chi(\tilde{\mu}_K^N(\omega) \in B^{\delta,N})\, P(\mathrm{d}\omega)$$

and

(5.13)
$$S_\omega^{\varepsilon,R}(K) = E_\omega(\exp\{K[\hat{F}(v_K) + \hat{P}(v_K)]\}\, \chi\{v_K \in M_\theta^{\varepsilon,R}\})$$

(5.14)
$$v_K = N^{-1} \sum_{l=1}^{N} \pi^2 \circ \vartheta^l(\mu_K^N).$$

Proof. Equation (5.11) follows from the uniform estimates in the proof of Lemma 17; (5.13) and (5.14) are obvious from the definitions (recall (3.34)). □

5b Large deviation analysis

After the preparations in Sect. 5.a we are now ready to do the large deviation analysis:

Lemma 19 *For every $\theta \in (0,1)$*

(5.15)
$$\lim_{N\to\infty} \lim_{\delta\to 0} \lim_{n\to\infty} \frac{1}{n} \log S^{\delta,N,\varepsilon,R}([\theta n]) = \theta \sup_{v \in M_{\theta,\beta\otimes\beta}^{\varepsilon,R}} [\hat{F}(v) + \hat{P}(v) - \hat{I}_{c,\beta^{\mathbb{Z}}}(v)]$$

with

(5.16)
$$M_{\theta,\beta\otimes\beta}^{\varepsilon,R} = \{v \in M_\theta^{\varepsilon,R} : \tilde{v} = \beta\otimes\beta\}$$

(5.17)
$$I_{c,\beta^{\mathbb{Z}}}(v) = -\sum_{i,j} v(i,j) \log((1-c)\, c^{j_1-1} \beta(j_2)) - \sup_{Q \in A_{v,\beta^{\mathbb{Z}}}} h(Q)$$

where $A_{v,\beta^{\mathbb{Z}}}$ and $h(Q)$ are defined prior to Theorem 2.

Proof. According to Lemma 18, $S^{\delta,N,\varepsilon,R}(K)$ is the expectation of an exponentional functional of μ_K^N over the double process $\{(X_k, b_k)\}$ restricted to a particular set enforced by the two indicators. Now, since this process is i.i.d. with finite state space $[1,R]\otimes \operatorname{supp}\beta$, the family (μ_K^N) with K running through the multiples of N, satisfies the large deviation principle on $\wp(([1,R]\otimes\operatorname{supp}\beta)^N)$ with rate function $\hat{I}_{c,\beta}^N(\mu)$ equal to the relative entropy of μ with respect to $\pi^N Q_{c,\beta}$, where $Q_{c,\beta}$ is the i.i.d. process with one dimensional marginal $(1-c)\, c^{j_1-1} \beta(j_2)$ and

π^N denotes the N-dimensional marginal (Ellis [9], Theorem VIII.2.1). As in the proof of Lemma 9 we can apply Varadhan's Theorem to (5.12) and obtain

(5.18)
$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log S^{\delta,N,\varepsilon,R}([\theta n]) &= \theta \sup_{\mu\in A^{\delta,N,\varepsilon,R}}\left[\hat F\left(N^{-1}\sum_{l=1}^{N}\pi^2\circ\vartheta^l(\mu)\right)\right.\\ &\qquad \left. + \hat P\left(N^{-1}\sum_{l=1}^{N}\pi^2\circ\vartheta^l(\mu)\right) - N^{-1}\hat I^N_{c,\beta}(\mu)\right]\end{aligned}$$

where

$$A^{\delta,N,\varepsilon,R} = \left\{\mu\in\wp(([1,R]\otimes\operatorname{supp}\beta)^N) \colon N^{-1}\sum_{l=1}^{N}\pi^2\circ\vartheta^l(\mu)\in M_\theta^{\varepsilon,R},\ \tilde\mu\in B^{\delta,N}\right\}.$$

This is a somewhat baroque expression, but relief is at hand. Observe that all objects on the r.h.s. of (5.18) are invariant under ϑ and therefore

(5.19)
$$\text{r.h.s.}(5.18) = \theta \sup_{\mu\in A_\vartheta^{\delta,N,\varepsilon,R}}[\hat F(\pi^2\mu) + \hat P(\pi^2\mu) - N^{-1}\hat I^N_{c,\beta}(\mu)]$$

where

$$A_\vartheta^{\delta,N,\varepsilon,R} = \{\mu\in\wp(([1,R]\otimes\operatorname{supp}\beta)^N)\colon \vartheta\mu=\mu,\ \pi^2\mu\in M_\theta^{\varepsilon,R},\ \tilde\mu\in B^{\delta,N}\}.$$

Let first $\delta\to 0$ to obtain, using the continuity of $\hat F(\cdot)$, $\hat P(\cdot)$ and $\hat I^N_{c,\beta}(\cdot)$ on $A_\vartheta^{\delta,N,\varepsilon,R}$, that

(5.20)
$$\lim_{\delta\to 0}\text{r.h.s.}(5.19) = \theta\sup_{\mu\in A_\vartheta^{N,\varepsilon,R}}[\hat F(\pi^2\mu)+\hat P(\pi^2\mu) - N^{-1}\hat I^N_{c,\beta}(\mu)],$$

where

$$\begin{aligned}A_\vartheta^{N,\varepsilon,R} &= \bigcap_{\delta>0} A_\vartheta^{\delta,N,\varepsilon,R}\\ &= \{\mu\in\wp(([1,R]\otimes\operatorname{supp}\beta)^N)\colon \vartheta\mu=\mu,\ \pi^2\mu\in M_\theta^{\varepsilon,R},\ \tilde\mu=\beta^N\}.\end{aligned}$$

Next let $N\to\infty$ to obtain

(5.21)
$$\lim_{N\to\infty}\text{r.h.s.}(5.20) = \text{r.h.s.}(5.15)$$

with

$$\begin{aligned}\hat I_{c,\beta^{\mathbb Z}}(v) &= \inf_{Q\in A_{v,\beta^{\mathbb Z}}}\hat I(Q\mid Q_{c,\beta})\\ \hat I(Q\mid Q_{c,\beta}) &= \lim_{N\to\infty}N^{-1}\hat I^N_{c,\beta}(\pi^N Q)\\ &= -\sum_{i,j}v(i,j)\log((1-c)\,c^{j_1-1}\beta(j_2)) - h(Q)\end{aligned}$$

(Ellis [9], Theorem IX.2.3 and p. 24). Combine (5.18–21). □

Finally it remains to show that

Lemma 20 *For every* $\theta\in(0,1)$

(5.22)
$$\begin{aligned}\sup_{R<\infty}\inf_{\varepsilon>0}\sup_{v\in M^{\varepsilon,R}_{\theta,\beta\otimes\beta}}[\hat F(v)+\hat P(v)-\hat I_{c,\beta^{\mathbb Z}}(v)] = \sup_{v\in M_{\theta,\beta\otimes\beta}}[\hat F(v)+\hat P(v)-\hat I_{c,\beta^{\mathbb Z}}(v)].\end{aligned}$$

Proof. This is a straightforward modification of the estimate in the proof of Lemma 10. Use that $\hat{I}_{c,\beta^{\mathbb{Z}}}(v)$ is lower semi continuous. □

The proof of Proposition 5 is now complete: combine Lemmas 15–20, recall the terms containing c in (3.18) which cancel against $\sum_{i,j} v(i,j)\log((1-c)\,c^{j_1-1})$ in (5.17), absorb $\hat{P}(v)$ into the remaining term $\sum_{i,j} v(i,j)\log\beta(j_2)$ in (5.17), to get the final result that $J^{\lambda}=$ r.h.s. (1.16). □

In Sects. 4 and 5 we have assumed that β has finite support. The proof for general distributions β (satisfying (0.15)) is very easy: approximate β by a discrete distribution β_0 with finite support such that $\|\beta-\beta_0\|\leqq\varepsilon$. In the exponents of (1.15) and (1.16) this introduces an error of at most $n\varepsilon\log(M/m)$ with M and m the maximal resp. minimal value in supp β. Let $n\to\infty$ followed by $\varepsilon\to 0$. Therefore (1.15) and (1.16) carry over immediately. □

As at the end of Sect. 3e one easily checks (1.11) and the analogue of (2.12).

6 Proof of Corollaries 1, 2 and 3

In this section we investigate the two variational formulas for $\rho(h)$ and $\lambda(h)$ given in Theorems 1 and 2. Section 6.a deals with $\rho(h)$, Sect. 6b with $\lambda(h)$. Both sections contain a proposition in which, besides proving Corollary 1 resp. 2, we condense several important facts about the variational formula, in order to convey a broader picture of the underlying structure. For instance, we obtain information about $\theta^*=\theta^*(h)$ and $v^*=v^*(h)$, the values where the suprema over θ and v are attained. As explained in Sect. 0c, $\theta^*(h)$ plays the role of effective drift of the typical path of descent. In particular, we shall see that $\theta^*(h)=0$ for h below a critical value, a property which we have called localization because it says that a typical particle at time n is within $o(n)$ of its ancestor. We also formulate and prove two assertions showing that certain interesting qualitative features of the phase diagram are direct consequences of some deeper but rather technical properties of the variational formulas. These properties will be addressed in Baillon et al. [1]. In Sect. 6.c we compare $\rho(h)$ and $\lambda(h)$, prove Corollary 3, and explain why $\rho(h)$ and $\lambda(h)$ can be equal for certain values of h and different for others.

6a $\rho(h)$

In Sect. 1.c we found that

(6.1) $$\rho(h)=\sup_{\theta\in[0,1]}(J^{\rho}(\theta)-I_h(\theta))$$

where for $\theta\in(0,1)$

(6.2) $$J^{\rho}(\theta)=\theta\sup_{v\in M_\theta}[\langle f\circ a, v\rangle-I_\theta(v)]$$

and for $\theta=0$ and $\theta=1$

(6.3) $$J^{\rho}(0)=\log M$$
$$J^{\rho}(1)=f(1).$$

It will be expedient to rewrite (6.1) and (6.2) in a form that is more suitable for manipulation. Define

$$J_h^\rho(\theta) = J^\rho(\theta) - I_h(\theta) \tag{6.4}$$

so that

$$\rho(h) = \sup_{\theta\in[0,1]} J_h^\rho(\theta). \tag{6.5}$$

The following relation will turn out to be very useful

$$\theta^{-1}[J_h^\rho(0) - J_h^\rho(\theta)] = -\frac{1}{2}\log\left(\frac{1+h}{1-h}\right) + \inf_{v\in M_\theta}[\langle g\circ a, v\rangle + I_0(v)]. \tag{6.6}$$

This relation is obtained by recalling $f(i) = i\log M - g(i)$ in (2.1) and by substituting into (6.2) the identity

$$\begin{aligned} I_0(v) - I_\theta(v) &= \sum_{i,j} v(i,j)\log(P_\theta(i,j)/P_0(i,j)) \\ &= \frac{1+\theta}{2\theta}\log(1+\theta) + \frac{1-\theta}{2\theta}\log(1-\theta) \\ &= \frac{1}{2}\log\left(\frac{1+h}{1-h}\right) - \theta^{-1}[I_h(0) - I_h(\theta)]. \end{aligned}$$

The second equality uses (3.3) plus the observation that $v\in M_\theta$ implies $\sum_{i,j} i\,v(i,j) = (1+\theta)/2\theta$ and $\sum_{i,j}(j-1)\,v(i,j) = (1-\theta)/2\theta$.

Proposition 6 *Assume that β has positive variance.*
(1) *$J_h^\rho(\cdot)$ is continuous and concave on $[0,1]$ for every h.*
(2) *The critical value for localization is*

$$h_1 = \frac{G^2-1}{G^2+1}$$

with

$$\begin{aligned} G &= \exp\left(\inf_{v\in\tilde M}[\langle g\circ a, v\rangle + I_0(v)]\right) \\ \tilde M &= \bigcup_{\theta>0} M_\theta \\ &= \{v\in\wp(\mathbb{N}^2): \sum_j v(i,j) = \sum_j v(j,i) \text{ for all } i\in\mathbb{N}, \langle a, v\rangle < \infty\}. \end{aligned}$$

Simple bounds on G are

$$\begin{aligned} &1 < M/A \leqq G \leqq 2M/A < \infty \\ &A = \int b\,\beta(\mathrm{d}b), \end{aligned}$$

so that $h_1\in(0,1)$.

(3) *The effective drift satisfies*

$$\theta^*(h)=0 \quad \textit{for } h<h_1$$
$$\theta^*(h)>0 \quad \textit{for } h>h_1.$$

(4) $\rho(\cdot)$ *is continuous on* $[0,1]$ *and*

$$\rho(h)=J_h^\rho(0)=\log[M(1-h^2)^{1/2}] \quad \textit{for } h\leqq h_1$$
$$\rho(h)>J_h^\rho(0) \quad \textit{for } h>h_1.$$

(5) *If* $\log M>0>f(1)$, *then there exists* $h_2\in(0,1)$ *where* $\rho(\cdot)$ *changes sign.*

Assertion A *If the supremum in* (6.2) *is attained in* $M_\theta\cap\{v>0\}$ *for every* $\theta\in(0,1)$, *then*

(*a*) $J_h^\rho(\cdot)$ *is strictly concave on* $[0,1]$ *for every* h;

(*b*) $\theta^*(h)$ *and* $v^*(h)$ *are unique for all* h.

If $J^\rho(\cdot)$ *is strictly decreasing on* $[0,1]$, *then*

(*c*) $0<\theta^*(h)<h$ *for all* $h_1<h<1$;

(*d*) $\theta^*(\cdot)$ *is continuous and strictly increasing on* $(h_1,1]$;

(*e*) $\rho(\cdot)$ *is strictly decreasing on* $[0,1]$ *and the critical value* h_2 *is unique.*

To verify the assumptions in Assertion A requires functional analytic techniques of some depth. We defer this question to a forthcoming paper: Baillon et al. [1]. For a simpler model where these properties are confirmed see Baillon et al. [0].

Remark. In [1] it is shown that the situation is slightly more complicated. The supposition only holds for $\theta\in(\theta_c,1)$ with some $\theta_c\in(0,1)$. This in consequence means that instead of (*a*) and (*b*) we have

(*a'*) $J_h^\rho(\cdot)$ *is linear on* $(0,\theta_c)$, *strictly concave on* $[\theta_c,1]$

(*b'*) $\theta^*(h)$ *exists and is unique for* $h\neq h_1$, $v^*(h)$ *exists and is unique for* $h>h_1$.

We shall not elaborate on this fact in this paper.

Proof of Proposition 6. (1) Since $I_h(\cdot)$ is continuous on $[0,1]$ we need to show that $J^\rho(\cdot)$ is continuous on $[0,1]$. The continuity on $(0,1)$ follows immediately from the observation in the last paragraph of Sect. 3.e. To prove continuity at 0 and 1, substitute $f(i)=i\log M-g(i)$ into (6.2) and use (6.3) to obtain, similarly as in (6.6),

$$\theta^{-1}[J^\rho(0)-J^\rho(\theta)]=\inf_{v\in M_\theta}[\langle g\circ a,v\rangle+I_\theta(v)]. \tag{6.7}$$

Because g is increasing and concave (see (2.1))

$$g(1)\leqq\theta^{-1}[J^\rho(0)-J^\rho(\theta)]\leqq g(\theta^{-1}),$$

where the second inequality follows from $\langle g\circ a,v\rangle\leqq g(\langle a,v\rangle)=g(\theta^{-1})(v\in M_\theta)$, together with the fact that $I_\theta(v)$ attains its minimal value zero at $v(i,j)=\pi_\theta(i)P_\theta(i,j)$ (see (3.3) and (3.4)), which is an element of M_θ. Now use that g is sublinear (see (2.1)) to get continuity of $J^\rho(\cdot)$ at $\theta=0$. Since g is continuous we also get continuity at $\theta=1$ via (6.3).

To prove concavity of $J_h^\rho(\cdot)$ we make the change of variable $\tau=\theta^{-1}$ and introduce

(6.8) $$K(\tau)=\tau[J_h^\rho(0)-J_h^\rho(\tau^{-1})]+\frac{1}{2}\log\left(\frac{1+h}{1-h}\right).$$

From (6.6) we have

(6.9) $$K(\tau)=\inf_{v\in M_{\tau^{-1}}}[\langle g\circ a, v\rangle+I_0(v)].$$

Below we shall show that $K(\tau)$ is convex in τ. This implies that $\tau^{-1}K(\tau)$ is convex in τ^{-1}. Hence the claim follows from (6.8).

Fix $\varepsilon>0$ and τ_1 and τ_2. Pick two measures v_1 and v_2 that are ε-optimal

$$K(\tau_1)\geqq -\varepsilon+[\langle g\circ a, v_1\rangle+I_0(v_1)] \quad (v_1\in M_{\tau_1^{-1}})$$
$$K(\tau_2)\geqq -\varepsilon+[\langle g\circ a, v_2\rangle+I_0(v_2)] \quad (v_2\in M_{\tau_2^{-1}}).$$

Since for every $0\leqq\alpha\leqq 1$

$$\langle a, \alpha v_1+(1-\alpha)v_2\rangle=\alpha\langle a, v_1\rangle+(1-\alpha)\langle a, v_2\rangle=\alpha\tau_1+(1-\alpha)\tau_2,$$

we have

$$\alpha v_1+(1-\alpha)v_2\in M_{[\alpha\tau_1+(1-\alpha)\tau_2]^{-1}}$$

and therefore from (6.9)

$$K(\alpha\tau_1+(1-\alpha)\tau_2)\leqq[\langle g\circ a, \alpha v_1+(1-\alpha)v_2\rangle+I_0(\alpha v_1+(1-\alpha)v_2)].$$

Next use that $I_0(v)$ is convex in v (see Ellis [9] p. 19)

$$I_0(\alpha v_1+(1-\alpha)v_2)\leqq\alpha I_0(v_1)+(1-\alpha)I_0(v_2)$$

to obtain

$$K(\alpha\tau_1+(1-\alpha)\tau_2)\leqq\varepsilon+\alpha K(\tau_1)+(1-\alpha)K(\tau_2).$$

Since ε can be made arbitrarily small we have finished the proof of (1).

(2) The continuity and concavity of $J_h^\rho(\cdot)$ implies that two situations are possible:

Case 1. $\theta^*=0$
Case 2. $\theta^*>0$.

The first case occurs when $J_h^\rho(\cdot)$ has a strictly negative slope at $\theta=0$. The concavity then implies that it is strictly decreasing on $[0,1]$ and therefore attains its maximum in (6.5) at the boundary $\theta=0$. The second case occurs when the slope at $\theta=0$ is strictly positive and the maximum is attained either in the interior $0<\theta<1$ or at the boundary $\theta=1$. We shall show using (6.6) that

$$\lim_{\theta\to 0}\theta^{-1}[J_h^\rho(0)-J_h^\rho(\theta)]=-\frac{1}{2}\log\left(\frac{1+h}{1-h}\right)+\inf_{v\in\tilde M}[\langle g\circ a, v\rangle+I_0(v)].$$

From this the critical value h_1 will follow as claimed.

Let us abbreviate

$$L(v)=\langle g\circ a, v\rangle + I_0(v).$$

Since $M_\theta \subset \tilde{M}$ for every $\theta>0$, we see from (6.6) that it suffices to show that for every $v\in\tilde{M}$ we can find a sequence $(v_\theta)_{\theta>0}$ with $v_\theta\in M_\theta$ such that

$$\limsup_{\theta\to 0} L(v_\theta)\leqq L(v). \tag{6.10}$$

Before we construct such a (v_θ), let us first observe that we may consider v with $v(i,j)=v(j,i)$. Indeed, $g(i+j-1)$ and $P_0(i,j)$ are both symmetric in i and j (see (3.3)) and hence the convexity of $I_0(v)$ implies

$$L\left(\frac{v+\tilde{v}}{2}\right)\leqq L(v)=L(\tilde{v}) \quad \text{where } \tilde{v}(i,j)=v(j,i),$$

so that the infimum in (6.6) may be restricted to the symmetric measures.

Therefore let v be a symmetric measure in $\tilde{M}$. Pick any pair (i,j) such that $v(i,j)=v(j,i)>0$ and pick a sequence $(\varepsilon_\theta)_{\theta>0}$ of positive reals tending to zero. For θ sufficiently small define v_θ by

$$v_\theta=v-\varepsilon_\theta(\delta_{(i,j)}+\delta_{(j,i)})+\varepsilon_\theta(\delta_{(i+k_\theta,\,j+k_\theta)}+\delta_{(j+k_\theta,\,i+k_\theta)})$$

where k_θ is determined by

$$\theta^{-1}=\langle a, v_\theta\rangle=\langle a, v\rangle+4\varepsilon_\theta\, k_\theta \tag{6.11}$$

in order to ensure that $v_\theta\in M_\theta$ (choose ε_θ such that k_θ is integer). Now observe that

$$\begin{aligned}L(v_\theta)-L(v)=&\,2\varepsilon_\theta[g(i+j+2k_\theta-1)-g(i+j-1)]\\ &-2\varepsilon_\theta\log[P_0(i+k_\theta,j+k_\theta)/P_0(i,j)]+H(\varepsilon_\theta)\end{aligned}$$

where $H(\varepsilon_\theta)$ is a sum of entropy terms tending to zero as ε_θ tends to zero. Use the concavity of g and the binomial inequality (3.27) to get

$$L(v_\theta)-L(v)\leqq 2\varepsilon_\theta\, g(2k_\theta)-2\varepsilon_\theta\log\left[\binom{2k_\theta}{k_\theta}2^{-2k_\theta}\right]+H(\varepsilon_\theta).$$

Next use Stirling's formula and (6.11) to see that the r.h.s. tends to zero when

$$\varepsilon_\theta\, g\left(\frac{1}{\theta\varepsilon_\theta}\right)\to 0$$

$$\varepsilon_\theta \log\left(\frac{1}{\theta\varepsilon_\theta}\right)\to 0.$$

But both g and log are sublinear and so this can be achieved by letting ε_θ tend to zero sufficiently fast with θ. This completes the proof of (6.10).

The lower bound on G is $\exp(g(1))=M/A$ (use (0.17)), because g is increasing and $I_0(v)\geqq 0$ for all v. The upper bound is obtained by substituting $v=\delta_{(1,1)}$ and noting that $P_0(1,1)=1/2$. This finishes the proof of (2).

(3) The claim follows from (6.5) and the remarks at the beginning of the proof of (2).

(4) Use (3), (6.3–5) and $I_h(0)=\log[(1-h^2)^{-1/2}]$. The continuity of $\rho(\cdot)$ follows from (6.5) because $I_h(\theta)$ is continuous in both h and θ and $J^\rho(\theta)$ is continuous in θ.

(5) Use the continuity of $\rho(\cdot)$ in conjunction with (0.18). □

Proof of Assertion A. Starting from the first assumption we can repeat the proof that $K(\tau)$ is convex in τ starting from optimality instead of ε-optimality (for the measures ν_1 and ν_2 in the proof of (1) of Proposition 6.) Accordingly, using that $I_0(\nu)$ is strictly convex on the subset of $\nu\in\mathscr{P}(\mathbb{N}^2)$ with $\nu>0$, we get that $K(\tau)$ is strictly convex in τ and hence that $\tau^{-1}K(\tau)$ is strictly convex in τ^{-1}. This proves (*a*) via (6.8). Of course, (*b*) follows from (*a*), the first assumption and the strict convexity of $I_\theta(\nu)$ in (6.2).

The second assumption implies (*c*) via (6.4) and (6.5), because the slope of $I_h(\cdot)$ changes sign at h. Next note that

$$\frac{\partial}{\partial h} I_h(\theta)=\frac{h-\theta}{1-h^2}.$$

With (*c*) this gives (*d*) because $(\partial/\partial h)\, I_h(\theta)$ is decreasing in θ, and (*e*), because $(\partial/\partial h)\, I_h(\theta)>0$ for $h>\theta$. □

6b $\lambda(h)$

Most of this section follows the same type of reasoning as in Sect. 6.a. In Sect. 1.c we found that

(6.12) $$\lambda(h)=\sup_{\theta\in[0,1]}(J^\lambda(\theta)-I_h(\theta))$$

where for $\theta\in(0,1)$

(6.13) $$J^\lambda(\theta)=\theta\sup_{\nu\in M_{\theta,\beta\otimes\beta}}[\langle \hat{f}\hat{a},\nu\rangle-I_{\theta,\beta^{\mathbb{Z}}}(\nu)]$$

and for $\theta=0$ and $\theta=1$

(6.14) $$J^\lambda(0)=\log M$$
$$J^\lambda(1)=\int\log b\,\beta(\mathrm{d}b).$$

Now, just as in (6.4–6) define

(6.15) $$J_h^\lambda(\theta)=J^\lambda(\theta)-I_h(\theta)$$

so that

(6.16) $$\lambda(h)=\sup_{\theta\in[0,1]}J_h^\lambda(\theta)$$

and

$$(6.17) \qquad \theta^{-1}[J_h^\lambda(0)-J_h^\lambda(\theta)] = -\frac{1}{2}\log\left(\frac{1+h}{1-h}\right) + \inf_{v\in M_{\theta,\beta\otimes\beta}} [\langle \hat{g}\hat{a}, v\rangle + I_{0,\beta^{\mathbf{z}}}(v)]$$

with

$$(6.18) \qquad \hat{g}(i,j) = \log(M/j_2) = \log M - \hat{f}(i,j).$$

Proposition 7 *Assume that β has positive variance.*
(1) *$J_h^\lambda(\cdot)$ is continuous and concave on $[0,1]$ for every h.*
(2) *The critical value for localization is*

$$h_3 = \frac{\hat{G}^2-1}{\hat{G}^2+1}$$

with

$$\hat{G} = \exp(\inf_{v\in M_{\beta\otimes\beta}} [\langle \hat{g}\hat{a}, v\rangle + I_{0,\beta^{\mathbf{z}}}(v)])$$

$$\tilde{M}_{\beta\otimes\beta} = \bigcup_{\theta>0} M_{\theta,\beta\otimes\beta}$$
$$= \{v\in \wp((N\otimes \operatorname{supp}\beta)^2): \sum_j v(i,j) = \sum_j v(j,j), \langle a, v\rangle < \infty, \tilde{v} = \beta\otimes\beta\}.$$

Simple bounds on $\hat{G}$ are

$$1 < M/B \leqq \hat{G} \leqq 2M/B < \infty$$
$$B = \exp(\textstyle\int \log b\,\beta(\mathrm{d}b)),$$

so that $h_3\in(0,1)$.
(3) *The effective drift satisfies*

$$\theta^*(h) = 0 \quad \text{for } h < h_3$$
$$\theta^*(h) > 0 \quad \text{for } h > h_3.$$

(4) *$\lambda(\cdot)$ is continuous on $[0,1]$ and*

$$\lambda(h) = J_h^\lambda(0) = \log[M(1-h^2)^{1/2}] \quad \text{for } h\leqq h_3$$
$$\lambda(h) > J_h^\lambda(0) \quad \text{for } h > h_3.$$

(5) *If $\log M > 0 > \int \log b\,\beta(\mathrm{d}b)$, then there exists $h_4\in(0,1)$ where $\lambda(\cdot)$ changes sign.*

Assertion B *If for every $\theta\in(0,1)$ the supremum in (6.13) is attained in $M_{\theta,\beta\otimes\beta}$ and $I_{\theta,\beta^{\mathbf{z}}}(\cdot)$ is strictly convex, then*
(a) *$J_h^\lambda(\cdot)$ is strictly concave on $[0,1]$ for every h;*
(b) *$\theta^*(h)$ and $v^*(h)$ are unique for all h.*
If $J^\lambda(\cdot)$ is strictly decreasing on $[0,1]$, then
(c) *$0 < \theta^*(h) < h$ for all $h_3 < h < 1$;*
(d) *$\theta^*(\cdot)$ is continuous and strictly increasing on $(h_3, 1]$;*
(e) *$\lambda(\cdot)$ is strictly decreasing on $[0,1]$ and the critical value h_4 is unique.*

Proof of Proposition 7. Most of the proof is a copy of the proof of Proposition 6 and therefore is left to the reader. The only somewhat tricky point is the continuity of $J^\lambda(\cdot)$ at $\theta=0$, which we shall now prove. In analogy with (6.7)

$$\theta^{-1}[J^\lambda(0)-J^\lambda(\theta)]=\inf_{\nu\in M_{\theta,\beta\otimes\beta}}[\langle\hat{g}\,\hat{a},\nu\rangle+I_{\theta,\beta^{\mathbb{Z}}}(\nu)]. \tag{6.19}$$

Therefore it suffices to find a sequence $(\nu_\theta)_{\theta>0}$ with $\nu_\theta\in M_{\theta,\beta\otimes\beta}$ such that

$$\lim_{\theta\to 0}\theta[\langle\hat{g}\,\hat{a},\nu_\theta\rangle+I_{\theta,\beta^{\mathbb{Z}}}(\nu_\theta)]=0. \tag{6.20}$$

To that end, for every $\theta>0$ pick $\xi_\theta:(\operatorname{supp}\beta)^2\to(0,1]$ satisfying

$$\sum_{i_2,j_2}\beta(i_2)\,\beta(j_2)\,\xi_\theta^{-1}(i_2,j_2)=\theta^{-1} \tag{6.21}$$

$$\xi_\theta(i_2,j_2)=\xi_\theta(j_2,i_2)$$

and choose

$$\nu_\theta(i,j)=\pi_{\xi_\theta(i_2,j_2)}(i_1)\,P_{\xi_\theta(i_2,j_2)}(i_1,j_1)\,\beta(i_2)\,\beta(j_2). \tag{6.22}$$

Using (3.3) and (3.4) one easily checks that $\nu_\theta\in M_{\theta,\beta\otimes\beta}$, where (6.21) is needed to take care of the requirements $\langle\hat{a},\nu_\theta\rangle=\theta^{-1}$ and $\sum_j\nu_\theta(i,j)=\sum_j\nu_\theta(j,i)$. Use the trivial inequality

$$I_{\theta,\beta^{\mathbb{Z}}}(\nu)\leqq-\sum_{i,j}\nu(i,j)\log(P_\theta(i_1,j_1)\,\beta(j_2))$$

and substitute (6.22) into (6.20) to get the upper bound for the r.h.s. of (6.19):

$$\theta\sum_{i_2,j_2}\beta(i_2)\,\beta(j_2)[\xi_\theta^{-1}(j_2)\log(M/j_2)+H(\theta,\xi_\theta(i_2,j_2))] \tag{6.23}$$
$$-\theta\sum_{j_2}\beta(j_2)\log\beta(j_2)$$

with

$$H(\theta,\xi)=-\sum_{i_1,j_1}\pi_\xi(i_1)\,P_\xi(i_1,j_1)\log P_\theta(i_1,j_1). \tag{6.24}$$

One should think of (6.22) as a *local drift strategy*. That is, ν_θ corresponds to a path of descent that behaves like the random walk but adapts its drift locally to the value of the medium: drift $\xi(i_2,j_2)$ on the pair (i_2,j_2). (This is not the optimal strategy). By clever choices for ξ_θ one can now get lower bounds for $J^\lambda(\theta)$, i.e. upper bounds in (6.19) and (6.20). We shall be happy with the crude choice

$$\xi_\theta(i_2,j_2)=1\quad\text{unless } i_2=j_2=M$$
$$\xi_\theta^{-1}(M,M)=1+\frac{\theta^{-1}-1}{\beta^2(M)}.$$

One easily checks that the r.h.s. of (6.23) tends to zero ((3.3–4) and supp β bounded away from 0) and hence we have proved (6.20) and thus the continuity of $J^{\lambda}(\theta)$ at $\theta=0$. If $\operatorname{supp}\beta$ is not finite then possibly $\beta(M)=0$. In that case adapt the choice of ξ_θ in a straightforward way.

The lower bound on $\hat{G}$ is trivial because $\langle \hat{g}\hat{a}, \nu\rangle \geqq \langle \hat{g}, \nu\rangle \geqq \sum_{j_2} \beta(j_2)\log(M/j_2)$ and $I_{0,\beta^z}(\nu)\geqq 0$ for all ν. The upper bound follows by substituting $\nu=(\delta_1\otimes\beta)^2$ for which $Q=(\delta_1\otimes\beta)^z$ maximises $h(Q)$.

Proof of Assertion B. By complete analogy with Assertion A. □

6c *Comparison of* $\rho(h)$ *and* $\lambda(h)$

In this section we show that the global growth rate is the supremum of local growth rates over media that are shift invariant processes. Let

$$A=\{\alpha\in\wp((\operatorname{supp}\beta)^z):\alpha \text{ shift invariant}\}$$
$$A_{\nu,\alpha}=\{Q\in\wp((N\otimes\operatorname{supp}\beta)^z): Q \text{ shift invariant}, \pi^2 Q=\nu, \tilde{Q}=\alpha\}$$
$$I_{\theta,\alpha}(\nu)=-\sum_{i,j}\nu(i,j)\log(P_\theta(i_1,j_1)\,\beta(j_2))-\sup_{Q\in A_{\nu,\alpha}} h(Q)\qquad(\nu\in A)$$

and define

$$J_\alpha^\lambda(\theta)=\theta\sup_{\nu\in M_{\theta,\pi^2\alpha}}[\langle\hat{f}\hat{a},\nu\rangle-I_{\theta,\alpha}(\nu)] \tag{6.25}$$

where $M_{\theta,\pi^2\alpha}$ is the set $M_{\theta,\beta\otimes\beta}$ with $\beta\otimes\beta$ replaced by $\pi^2\alpha$.

Proposition 8 *For every* $\theta\in(0,1)$

$$J^\lambda(\theta)=J^\lambda_{\beta^z}(\theta) \tag{6.26}$$

$$J^\rho(\theta)=\sup_{\alpha\in\mathscr{A}} J_\alpha^\lambda(\theta). \tag{6.27}$$

Observe that these relations imply Corollary 3 via (6.4–5) and (6.15–16).

Proof. (6.26) is just (6.13). To see that (6.27) produces (6.2) we argue as follows. Measures in $M_{\theta,\pi^2\alpha}$ have the property that their projection on the non-medium coordinates is in M_θ, the set over which runs the supremum in (6.2) for $J^\rho(\theta)$. Therefore define for $\mu\in M_\theta$

$$N_\mu=\{\nu\in\wp((\mathbb{N}\otimes\operatorname{supp}\beta)^2): \sum_j \nu(i,j)=\sum_j \nu(j,i), \sum_{i_2,j_2}\nu(i,j)=\mu(i_1,j_1)\}.$$

Now, by the contraction principle (Ellis [9] Theorem IX.3.3)

$$\sup_{Q\in A_\nu} h(Q)=-\sum_{i,j}\nu(i,j)\log\left(\frac{\nu(i,j)}{\bar{\nu}(i)}\right) \tag{6.28}$$

with

$$\begin{aligned} A_\nu &= \bigcup_{\alpha\in\mathscr{A}} A_{\nu,\alpha} \\ &= \{Q\in\wp((\mathbb{N}\otimes\operatorname{supp}\beta)^z): Q \text{ shift invariant}, \pi^2 Q=\nu\}. \end{aligned} \tag{6.29}$$

Since

$$\bigcup_{\alpha\in\mathscr{A}} M_{\theta,\pi^2\alpha} = \bigcup_{\mu\in M_\theta} N_\mu$$

it follows from (6.25) and (6.28–29) that we can write

$$\sup_{\alpha\in\mathscr{A}} J_\alpha^\lambda(\theta) = \theta \sup_{\mu\in M_\theta}\{\sup_{\nu\in N_\mu}[\langle \hat{f}\hat{a}, \nu\rangle - \tilde{I}_{\theta,\beta}(\nu)]\}, \tag{6.30}$$

with

$$\tilde{I}_{\theta,\beta}(\nu) = \sum_{i,j} \nu(i,j)\log\left(\frac{\nu(i,j)}{\bar{\nu}(i)\, P_\theta(i_1,j_2)\,\beta(j_2)}\right). \tag{6.31}$$

Thus (6.27) amounts to proving that for every $\mu\in M_\theta$

$$\sup_{\nu\in N_\mu}[\langle \hat{f}\hat{a}, \nu\rangle - \tilde{I}_{\theta,\beta}(\nu)] = \langle \hat{f}\circ a, \mu\rangle - I_\theta(\mu)$$

(recall (6.2)), or written out

$$\begin{aligned}\sup_{\nu\in N_\mu}&\left[\sum_{i,j}\nu(i,j)\log\gamma(i_1,j) - \sum_{i,j}\nu(i,j)\log\left(\frac{\nu(i,j)}{\bar{\nu}(i)}\right)\right]\\ &= \sum_{i_1,j_1}\mu(i_1,j_1)\log\{\sum_{j_2}\gamma(i_1,j)\} - \sum_{i_1,j_1}\mu(i_1,j_1)\log\left(\frac{\mu(i_1,j_1)}{\bar{\mu}(i_1)}\right)\end{aligned} \tag{6.32}$$

with the abbreviation

$$\gamma(i_1,j) = \beta(j_2)\, j_2^{i_1+j_1-1}$$

(the kernel $P_\theta(i_1,j_1)$ drops out). The reader can easily check (6.32) informally by using the technique of Lagrange multipliers. The supremum is attained at

$$\begin{aligned}\nu(i,j) &= \bar{\nu}(i)\, R(i,j)\\ \bar{\nu}(j) &= \sum_i \bar{\nu}(i)\, R(i,j)\end{aligned}$$

with transition kernel

$$R(i,j) = \frac{\mu(i_1,j_1)}{\bar{\mu}(i_1)}\,\frac{\gamma(i_1,j)}{\{\sum_{j_2}\gamma(i_1,j)\}}.$$

Substitution into (6.32) shows that the supremum does not depend on $\bar{\nu}(i)$, so that we need not solve for $\bar{\nu}(i)$. This is an informal proof because there is of course the technical problem that $\mathbb{N}\otimes\operatorname{supp}\beta$ is infinite. However, we saw in Lemmas 10 and 20 that our variational formulas are limits of variational formulas on finite state space and therefore the argument can be made rigorous. □

Assertion C *If for every $\theta\in(0,1)$ the supremum in* (6.27) *is attained at $\alpha^* = \alpha^*(\theta)\in\mathscr{A}$ with $\alpha^*\neq\beta^{\mathbb{Z}}$, then $\rho(h)>\lambda(h)$ for all $h>h_1$.*

Proof. Obvious from Proposition 8. Use (6.4–5) and (6.15–16) plus the fact that for $h>h_3$ both locally and globally $\theta^*>0$. We know from Corollaries 1 and 2 that $\rho(h)>\lambda(h)$ for $h_1<h\leqq h_3$. □

Acknowledgement. The work on this paper was supported by the Deutsche Forschungsgemeinschaft through the SFB 123 at the Universität Heidelberg and by the Royal Netherlands Academy of Arts and Sciences. FdH was a guest of the SFB 123 at the Universität Heidelberg during the Spring and Summer of 1989.

References

0. Baillon, J.-B., Clément, Ph., Greven, A., den Hollander, F.: A variational approach to branching random walk in random environment. Ann. Probab. (to appear)
1. Baillon, J.-B., Clément, Ph., Greven, A., den Hollander, F.: On a variational problem for an infinite particle system in a random medium (preprint, 1991)
2. Bramson, M., Durrett, R., Schonmann, R.H.: The contact process in a random environment. Ann. Probab. **19**, 960–983 (1990)
3. Chen, D., Liggett, T.M.: Finite reversible nearest particle systems in inhomogeneous and random environments (preprint, 1989)
4. Dawson, D., Fleischmann, K.: On spatially homogeneous branching processes in a random environment. Math. Nachr. **113**, 249–257 (1983)
5. Dawson, D., Fleischmann, K.: Critical dimension for a model of branching in a random medium, Z. Wahrscheinlichkeitstheor. Verw. Geb. **70**, 315–334 (1985)
6. Dawson, D., Fleischmann, K.: Critical branching in a highly fluctuating random medium. Probab. Theory Relat. Fields **90**, 241–274 (1991)
7. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations I–IV, Commun. Pure Appl. Math. **28**, 1–47 (1975); **28**, 279–301 (1975); **29**, 389–461 (1976); **36**, 183–212 (1983)
8. Eisele, Th., Lang, R.: Asymptotics for the Wiener Sausage with drift. Probab. Theory Relat. Fields **74**, 125–140 (1987)
9. Ellis, R.: Entropy, large deviations and statistical mechanics. Grundlehren der mathematischen Wissenschaften, vol. 271. Berlin Heidelberg New York: Springer 1985
10. Ferreira, I.M.: Cluster for the voter model in random environment and probability of survival for the biased voter model in random environment. Ph.D. thesis Cornell University 1988
11. Gärtner, J., Molchanov, S.A.: Parabolic problems for the Anderson model. I. Intermittency and related topics. Commun. Math. Phys. **132**, 613–655 (1990)
12. Greven, A.: The coupled branching process in random environment. Ann. Probab. **13**, 1133–1147 (1985)
13. Greven, A.: A class of infinite particle systems in random environment. In: Tautu, P. (ed.) (Lect. Notes Math. vol. 1212, pp. 145–164) Berlin Heidelberg New York: Springer 1986
14. Greven, A.: Symmetric exclusion on random sets and a related problem for random walks in random environment. Probab. Theory Relat. Fields **85**, 307–364 (1990)
15. Greven, A., den Hollander, F.: Population growth in a random media. I. Variational formula and Phase Diagram. II. Wave front propagation. J. Stat. Phys. **65**, (5/6) 1991
16. Kasteleyn, P.W.: Graph theory and crystal physics. In: Harary F (ed.) Graph theory and theoretical physics. London: Academic Press 1967
17. Liggett, T.M.: Spatially inhomogeneous contact processes. In: Proceedings of a Conference in Honor of Ted Harris. Boston: Birkhäuser
18. Ney, P., Nummelin, E.: Markov additive processes II. Large deviations. Ann. Probab. **15**, 593–609 (1987)
19. Stroock, D.W.: An introduction to the theory of large deviations. Berlin Heidelberg New York: Springer 1984
20. Varadhan, S.R.S.: Asymptotic probabilities and differential equations. Commun. Pure Appl. Math. **19**, 261–286 (1966)