Relative entropy in a variational study of branching random walk in random environment *

F. den Hollander

EURANDOM, P.O. Box 513,
5600 MB Eindhoven, The Netherlands

Hollander@eurandom.tue.nl

2nd July 2000

Abstract

We consider an infinite system of particles on the integer lattice that: (1) migrate to the right with a random delay, (2) branch along the way according to a random law depending on their position. The initial configuration has one particle at each site. With the help of large deviation theory, we compute the exponential growth rate of the average number of particles in a large box (= global growth rate) and the average number of particles at the origin (= local growth rate). Both these growth rates are expressed in terms of a variational problem. An analysis and comparison of these variational problems reveals various interesting phase transitions as a function of the underlying parameters. A key role in the calculations is played by the notion of *relative entropy*, which is used to compute the weight of the various growth strategies of the system and to identify the optimal growth strategy that determines the growth rate.

AMS 1991 *subject classifications.* 60F10, 60J15, 82B44.
Key words and phrases. Branching random walk in random environment, large deviations, relative entropy, variational problems, phase transitions.

1 Introduction

1.a. Motivation. The contributions in this volume deal with the different faces that entropy has in probability theory, ergodic theory, dynamical systems, information theory, statistical physics and thermodynamics. In the present contribution we focus on an application of *relative entropy* in probability theory and statistical physics, namely, to population growth in random media.

Over the past ten years or so, I have been working with Andreas Greven on a program in which we study random processes that interact either with themselves or with a random medium and we try to describe their evolution using large deviation theory, entropy techniques

*Paper presented at the "International Symposium on Entropy" hosted by the Max Planck Institute for Physics of Complex Systems, Dresden, Germany, 26–28 June 2000.

and variational calculus. We have so far looked at polymer chains, diffusions in random potentials, interacting diffusions, and population growth in random media (see [8], Part B, and references therein). In order to explain what the program is about, I have chosen the simplest example studied so far, namely, particles doing a one-sided migration on the integer lattice and branching according to a local offspring distribution that is chosen randomly and independently for different sites. This example is typical for the type of ideas and techniques that come up also in more complicated examples. The present paper is a review of the work in [1], [2], [4], [5], [6], [7].

We begin by formulating *Sanov's Theorem in large deviation theory*, which is the main conceptual tool in what follows (see [3], Section 3, and [8], Theorem II.36). Let $(Z_i)_{i\in\mathbb{N}}$ be i.i.d. random variables on a countable state space $\mathbb{S}$ with marginal probability law μ. For $n\in\mathbb{N}$, let

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i} \tag{1.1}$$

be the empirical measure associated with $Z_1,\dots,Z_n$. This is a random element of $\mathcal{P}(\mathbb{S})$, the set of probability measures on $\mathbb{S}$, which is endowed with the total variation distance. By the ergodic theorem, L_n tends to μ in $\mathcal{P}(\mathbb{S})$ as $n\to\infty$. Sanov's Theorem says that the family $(L_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle on $\mathcal{P}(\mathbb{S})$, i.e.,

$$\begin{aligned}
&\limsup_{n\to\infty}\tfrac{1}{n}\log\mu^{\mathbb{N}}(L_n\in C)\leq -\inf_{\nu\in C}H(\nu|\mu)\quad \forall C\subset\mathcal{P}(\mathbb{S})\text{ closed},\\
&\liminf_{n\to\infty}\tfrac{1}{n}\log\mu^{\mathbb{N}}(L_n\in O)\geq -\inf_{\nu\in O}H(\nu|\mu)\quad \forall O\subset\mathcal{P}(\mathbb{S})\text{ open},
\end{aligned}\tag{1.2}$$

where the rate function $H(\cdot|\mu)\colon\mathcal{P}(\mathbb{S})\to[0,\infty]$ is given by

$$\begin{aligned}
H(\nu|\mu) &= \sum_{s\in\mathbb{S}}\nu(s)\,\log\tfrac{\nu(s)}{\mu(s)}\\
&= \textit{relative entropy}\text{ of }\nu\text{ w.r.t. }\mu.
\end{aligned}\tag{1.3}$$

What (1.2) says is that for all $\nu\in\mathcal{P}(\mathbb{S})$ where H is continuous:

$$\mu^{\mathbb{N}}(L_n\approx\nu)=\exp\Big[-H(\nu|\mu)\,n+o(n)\Big],\qquad n\to\infty. \tag{1.4}$$

Thus, Sanov's Theorem gives us a precise description of the large deviation probabilities of L_n away from its typical value μ in terms of the rate function H. It is important to note that $H(\nu|\mu)\geq 0$ with equality if and only if $\nu=\mu$. Moreover, $\nu\mapsto H(\nu|\mu)$ is strictly convex, lower semi-continuous, and has compact level sets (see [8], Lemma II.39). We may think of $H(\nu|\mu)$ as a kind of distance between ν and μ: the larger this distance the more costly the large deviation.

1.b. Branching random walk in random environment. In Sections 1.b to 1.f we describe the work in [1], [4].

With each $x\in\mathbb{Z}$ is associated a random probability measure F_x on $\mathbb{N}_0=\mathbb{N}\cup\{0\}$, called the offspring distribution at site x. The sequence

$$F=\{F_x\}_{x\in\mathbb{Z}} \tag{1.5}$$

is i.i.d. with common distribution α. For fixed F, define a discrete-time Markov process $(\eta_n)_{n\in\mathbb{N}_0}$ on $\mathbb{N}_0^{\mathbb{Z}}$, with the interpretation

$$\begin{aligned}
\eta_n &= \{\eta_n(x)\}_{x\in\mathbb{Z}},\\
\eta_n(x) &= \text{number of particles at site } x \text{ at time } n,
\end{aligned}\tag{1.6}$$

by specifying its one step transition mechanism as follows. Start from initial state $\eta_0 \equiv 1$. Given the state η_n at time n:

(i) Each particle is independently replaced by a new generation. The size of a new generation descending from a particle at site x has distribution F_x. All particles branch independently.

(ii) Immediately after creation, each new particle decides to jump one lattice spacing to the right with probability h or stand still with probability $1-h$, with $h \in (0,1)$.

The resulting sequence of particle numbers after steps (i) and (ii) make up the state η_{n+1} at time $n+1$, etc.

The sequence F plays the role of a *random environment* and stays fixed during the evolution. The parameter h is the *drift of the migration*. We write $\mathbb{P}, \mathbb{E}$ and P, E to denote probability and expectation for the environment resp. the migration process.

Let b_x denote the average offspring at site x and let β denote the distribution of b_x induced by α. Write $\mathbb{B}$ to denote the support of β. We assume that

$$\begin{aligned} &\mathbb{B} \text{ is countable and not a singleton,} \\ &\mathbb{B} \text{ is bounded away from 0 and } \infty, \end{aligned} \tag{1.7}$$

and write

$$M = \text{ supremum of } \mathbb{B}. \tag{1.8}$$

1.c. Particle densities and growth rates. We will consider the following two particle densities:

$$\begin{aligned} d_n &= (\mathbb{E} \times E)(\eta_n(0)) &&= \textit{global} \text{ particle density at time } n, \\ \hat{d}_n(F) &= E(\eta_n(0) \mid F) &&= \textit{local} \text{ particle density at time } n. \end{aligned} \tag{1.9}$$

Here, note that for each n the sequence η_n is stationary and ergodic under $\mathbb{P} \times P$, so that by the ergodic theorem we have

$$d_n = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{x=-N}^{N} \eta_n(x) \qquad (\mathbb{P} \times P) - a.s., \tag{1.10}$$

showing that indeed d_n is the particle density in a large box. We will be interested in computing the exponential growth rates

$$\begin{aligned} \lambda &= \lim_{n \to \infty} \tfrac{1}{n} \log d_n, \\ \hat{\lambda}(F) &= \lim_{n \to \infty} \tfrac{1}{n} \log \hat{d}_n(F). \end{aligned} \tag{1.11}$$

These are identified in Theorems 1 and 2 below.

For $\theta \in [0,1]$, let B_θ and π_θ denote the Bernoulli distribution resp. the geometric distribution with parameter θ given by

$$\begin{aligned} &B_\theta(0) = 1-\theta,\ B_\theta(1) = \theta, \\ &\pi_\theta(i) = \theta(1-\theta)^{i-1},\ i \in \mathbb{N}. \end{aligned} \tag{1.12}$$

Theorem 1 [1] *For $h \in (0,1)$ and β as in (1.7):*

$$\lambda = \lambda(\beta, h) \tag{1.13}$$

with

$$\lambda(\beta,h) = \sup_{\theta\in(0,1]} \Big[\theta\Big\{\sup_{\nu\in\mathcal{M}_\theta}[f(\nu) - H(\nu|\pi_\theta)]\Big\} - H(B_\theta|B_h)\Big], \tag{1.14}$$

where

$$\begin{aligned} \mathcal{M}_\theta &= \Big\{\nu \in \mathcal{P}(\mathbb{N})\colon \textstyle\sum_{i\in\mathbb{N}} i\nu(i) = \frac{1}{\theta}\Big\}, \\ f(\nu) &= \textstyle\sum_{i\in\mathbb{N}} \nu(i) \log \sum_{j\in\mathbb{B}} \beta(j) j^i. \end{aligned} \tag{1.15}$$

Theorem 2 [4] *For $h \in (0,1)$ and β as in (1.7):*

$$\hat\lambda(F) = \hat\lambda(\beta,h) \qquad \mathbb{P} - a.s. \tag{1.16}$$

with

$$\hat\lambda(\beta,h) = \sup_{\theta\in(0,1]} \Big[\theta\Big\{\sup_{\nu\in\hat{\mathcal{M}}_{\theta,\beta}}[\hat f(\nu) - H(\nu|\pi_\theta\times\beta)]\Big\} - H(B_\theta|B_h)\Big], \tag{1.17}$$

where

$$\begin{aligned} \hat{\mathcal{M}}_{\theta,\beta} &= \Big\{\nu \in \mathcal{P}(\mathbb{N}\times\mathbb{B})\colon \textstyle\sum_{i\in\mathbb{N},j\in\mathbb{B}} i\nu(i,j) = \frac{1}{\theta},\ \sum_{i\in\mathbb{N}} \nu(i,j) = \beta(j)\ \forall j\in\mathbb{B}\Big\}, \\ \hat f(\nu) &= \textstyle\sum_{i\in\mathbb{N},j\in\mathbb{B}} \nu(i,j)\log j^i. \end{aligned} \tag{1.18}$$

Note that the variational problems in (1.14) and (1.17) are similar in structure, but different. In Section 1.e. we will explain where they come from.

The link between the global and the local growth rate is expressed by the following theorem.

Theorem 3 [1], [4] *For $h \in (0,1)$ and β as in (1.7):*

$$\lambda(\beta,h) = \sup_{\gamma\in\mathcal{P}(\mathbb{B})} [\hat\lambda(\gamma,h) - H(\gamma|\beta)]. \tag{1.19}$$

Also this relation will be explained in Section 1.e.

1.d. Solution of the variational problems. The variational problems in Theorems 1 and 2 can be solved explicitly. Indeed, they involve maximization of functionals containing relative entropy under linear constraints, which can be achieved with the help of exponential families of probability measures via the Gibbs-Jaynes principle cited in [3], Section 4.

Define, for $r \geq 0$,

$$\begin{aligned} G(r) &= \textstyle\sum_{j\in\mathbb{B}} \beta(j)\Big(\frac{e^{-r}[j/M]}{1-e^{-r}[j/M]}\Big), \\ \hat G(r) &= \exp\Big[\textstyle\sum_{j\in\mathbb{B}} \beta(j)\log\Big(\frac{e^{-r}[j/M]}{1-e^{-r}[j/M]}\Big)\Big]. \end{aligned} \tag{1.20}$$

Define

$$\begin{aligned} h_c &= \lim_{r\downarrow 0} \tfrac{1}{1+G(r)}, \\ \theta_c &= \lim_{r\downarrow 0} \tfrac{1}{-[\log G]'(r)}, \end{aligned} \tag{1.21}$$

and similarly for $\hat{G}$. Note that

$$-[\log G]'(r) > 1 + G(r) = -[\log \hat{G}]'(r) > 1 + \hat{G}(r) \qquad \forall r \geq 0, \tag{1.22}$$

implying that

$$\theta_c < \hat{\theta}_c = h_c < \hat{h}_c. \tag{1.23}$$

To express our solution in a compact form we need two more quantities, $r^* = r^*(\beta, h)$ and $\theta^* = \theta^*(\beta, h)$, defined as follows:

$$\begin{array}{ll} h \leq h_c: & r^* = 0, \\ h > h_c: & r^* \text{ is the unique solution of } h = \frac{1}{1+G(r)}, \\ & \\ h \leq h_c: & \theta^* = 0, \\ h > h_c: & \theta^* = \frac{1}{-[\log G]'(r^*)}, \end{array} \tag{1.24}$$

and similarly for $\hat{G}$.

Theorem 4 [1], [4] *(See Figs. 1 and 2.) Fix β subject to (1.7).*
(i) For $h \in (0,1)$:

$$\begin{aligned} \lambda(\beta, h) &= \log[M(1-h)] + r^*(\beta, h), \\ \hat{\lambda}(\beta, h) &= \log[M(1-h)] + \hat{r}^*(\beta, h). \end{aligned} \tag{1.25}$$

(ii) $h \mapsto \lambda(\beta, h)$ is continuous and strictly decreasing on $(0,1)$, analytic on $(0, h_c)$ and $(h_c, 1)$, while at the boundary points:

$$\lambda(\beta, 0) = \log M, \qquad \lambda(\beta, 1) = \log \sum_{j \in \mathbb{B}} \beta(j)\, j. \tag{1.26}$$

(iii) $h \mapsto \hat{\lambda}(\beta, h)$ is continuous and strictly decreasing on $(0,1)$, analytic on $(0, \hat{h}_c)$ and $(\hat{h}_c, 1)$, while at the boundary points:

$$\hat{\lambda}(\beta, 0) = \log M, \qquad \hat{\lambda}(\beta, 1) = \sum_{j \in \mathbb{B}} \beta(j)\, \log j. \tag{1.27}$$

(iv) $\lambda(\beta, h) \geq \hat{\lambda}(\beta, h)$ with strict inequality if and only if $h > h_c$.
(v) If $h_c > 0$, then

$$\frac{\partial}{\partial h}\lambda(\beta, h_c+) - \frac{\partial}{\partial h}\lambda(\beta, h_c-) = \frac{\theta_c}{h_c(1-h_c)}. \tag{1.28}$$

Similarly for $\hat{\lambda}(\beta, h), \hat{h}_c, \hat{\theta}_c$.
(vi) For $h \in (0,1)$, $\theta^*(\beta, h)$ and $\hat{\theta}^*(\beta, h)$ are the maximizers of the outer variational problem in (1.14) resp. (1.17).
(vii) On $(h_c, 1)$, $h \mapsto \theta^*(\beta, h)$ is strictly increasing, analytic and satisfies $\theta_c < \theta^*(\beta, h) < h$. Similarly for $\hat{\theta}^*(\beta, h), \hat{h}_c, \hat{\theta}_c$.

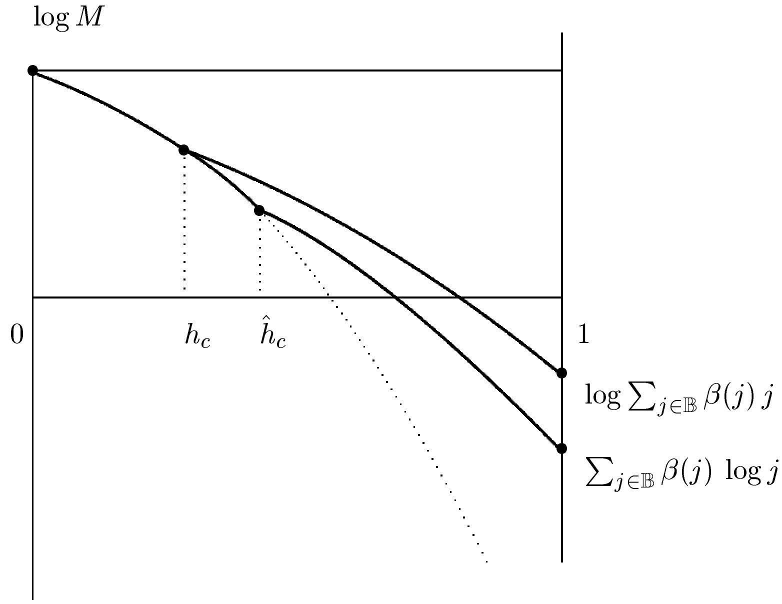

Fig. 1 Qualitative picture of $h \mapsto \lambda(\beta, h)$ (upper solid curve), $h \mapsto \hat{\lambda}(\beta, h)$ (lower solid curve) and $h \mapsto \log[M(1-h)]$ (dotted curve). The solid curves split off the dotted curve at h_c resp. $\hat{h}_c$.

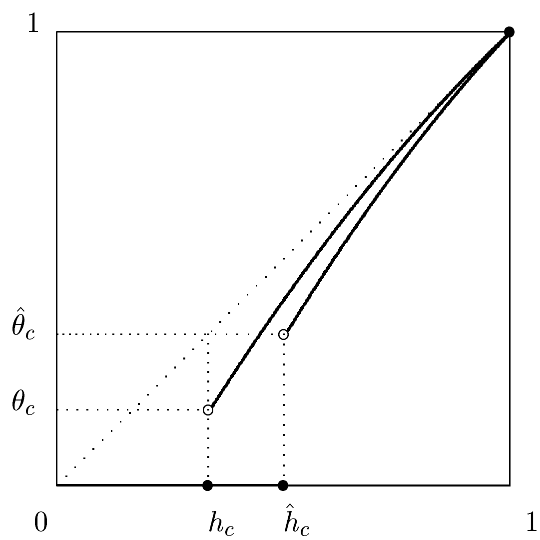

Fig. 2 Qualitative picture of $h \mapsto \theta^*(\beta, h)$ (upper solid curve) and $h \mapsto \hat{\theta}^*(\beta, h)$ (lower solid curve). The solid curves jump from 0 to θ_c and $\hat{\theta}_c$ at h_c resp. $\hat{h}_c$.

1.e. Interpretation of Theorems 1 and 2. What drives Theorems 1 and 2 is a close interplay between the migration process and the environment. The idea is that, for large n, *the population predominantly consists of those particles whose history happens to be best adapted to the environment*. Here, best adapted means that the particles have a *path of descent* that spends a lot of time on sites x where b_x is large, little time on sites x where b_x is small, and does so in a way that is not too unlikely. Indeed, such particles are part of a family that produces the most offspring and therefore dominates the population ("survival of the fittest").

Note that when we say population, we must distinguish between two types of population:

$$\begin{array}{l}\text{the } \textit{global} \text{ population } (= \text{the population in a large box}),\\ \text{the } \textit{local} \text{ population } (= \text{the population at the origin}).\end{array} \tag{1.29}$$

Two particles drawn randomly from these two populations will have different paths of descent.

We now explain where the two variational problems in Theorems 1 and 2 come from. We give only the main idea. The details will be explained in Section 2.

Let us first look at the global variational problem. Note that this variational problem has a two-layer structure: it consists of an outer and an inner variational problem. The reason for this is as follows. What happens is that the path of descent of a particle may assume any *empirical drift* θ and any *empirical local time law* ν, i.e., the empirical measure for the steps in the path resp. for the time spent at the sites that are visited successively. These are the two variables appearing in (1.14). The cost of adopting θ is $H(B_\theta|B_h)$, the cost of adopting ν given θ is $H(\nu|\pi_\theta)$. Conditioned on adopting ν, the path of descent produces offspring at rate $f(\nu)$ in (1.15), which is the growth factor on a site whose local time law is ν when averaged over the local environment at that site. Thus we see where (1.14) comes from: the growth rate is determined by the optimal choice for θ and ν in a competition between cost and gain. Note that the constraint in $\mathcal{M}_\theta$ in (1.15) says that the average local time at a site must be compatible with the drift, so we have a constrained variational problem.

For the local variational problem the structure is the same, except that the environment is fixed. In this case ν plays the role of the *empirical joint local time law and local environment law*, keeping track of which local time occurs on top of which local environment. The cost of adopting ν given θ is $H(\nu|\pi_\theta \times \beta)$, and conditioned on ν the path of descent produces offspring at rate $\hat{f}(\nu)$ in (1.18), which is the growth factor on a site whose joint local time law and local environment law are ν. This explains (1.17). Note that the set $\hat{M}_{\beta,\theta}$ in (1.18) has two constraints: the first again says that the average local time at a site must be compatible with the drift, the second says that the projection onto the environment coordinate must equal β.

The explanation of the link established in Theorem 3 is as follows. The global particle density is the average over the random environment of the local particle density (recall (1.9)). Effectively, this means that for the global growth rate the random environment participates in the competition between cost and gain, while for the local growth rate it does not. The cost for the environment of adopting an *empirical local environment law* γ is $H(\gamma|\beta)$. Conditioned on adopting γ, the growth rate is $\hat{\lambda}(\gamma, h)$. Thus, (1.19) says that the global growth rate is determined by the optimal choice for γ: most of the population in a large box comes from those stretches in the box where the environment is optimal.

It turns out that (1.14) and (1.17) have unique maximizers. In fact, it can be proved that the path of descent of a typical particle drawn from the global resp. the local population has an empirical drift and an empirical local time law resp. an empirical joint local time law and local environment law that converges to the unique maximizers of the variational problems. This is related to Csiszár's conditional limit theorem cited in [3], Section 5. Also (1.19) has a unique maximizer, and along the path of descent the empirical local environment law converges to this maximizer.

1.f. Phase transitions. Theorem 4 shows that our system exhibits various interesting phase transitions. We now interpret these phase transitions for the case where β, in addition to (1.7), has the following properties (already assumed in Figs. 1 and 2):

(I) $\log M > 0 > \log \sum_{j \in \mathbb{B}} \beta(j)\, j$: both solid curves in Fig. 1 cross zero.

(II) $\log[M(1-\hat{h}_c)] > 0$: both solid curves in Fig. 1 split off the dashed curve before crossing zero.

(III) $\sum_{j\in\mathbb{B}} \beta(j)(1-[j/m])^{-2} < \infty$: both solid curves in Fig. 2 jump.

(I) *Survival vs. extinction*: Fig. 1 shows that there is a critical value for h at which $\lambda(\beta,h)$ changes sign. Below this value the population grows (= global survival), above this value the population decays (= global extinction). The same interpretation applies for $\hat{\lambda}(\beta,h)$.

(II) *Clumping*: Fig. 1 shows that there is an intermediate range of h-values where $\lambda(\beta,h) > 0 > \hat{\lambda}(\beta,h)$, i.e., global survival but local extinction. This means that the particles are strongly clustering together (= clumping): the density of populated sites decays to zero, but the population on these sites grows so fast that the overall particle density still grows.

(III) *Localization vs. delocalization*: Fig. 2 shows that for $h \leq h_c$ we have $\theta^*(\beta,h) = 0$, meaning that the typical path of descent moves at sublinear speed (= global localization), while or $h > h_c$ we have $\theta^*(\beta,h) > 0$, meaning that the typical path of descent moves at linear speed (= global delocalization). At $h = h_c$ the speed makes a jump of size $\theta_c > 0$. The dashed curve in Fig. 1 corresponds to the strategy where the typical path of descent of a particle drawn from the global resp. the local population spends most of its time on sites where the environment is maximal. The cost of this strategy is $H(B_0|B_h) = \log(1/(1-h))$, the gain is $\log M$. Below h_c this strategy is optimal, above h_c it is not. The random environment has the tendency to slow down the path where it is large and to speed up the path where it is small. Since $\theta^*(\beta,h) < h$, as shown in Fig. 2, the overall effect of the random environment apparently is to slow down the path. The same interetation applies for $\hat{\theta}^*(\beta,h), \hat{h}_c, \hat{\theta}_c$.

1.g. Outline. The remainder of this paper is organized as follows. In Section 2 we sketch the derivation of Theorems 1, 2 and 3 given in [1], [4], which is based on Sanov's Theorem in Section 1.a. In Section 3 we discuss two extensions: one described in [5] where the initial state is a step function and one described in [2], [6], [7] where the migration runs both forward and backward. In Section 4 we close by formulating some open problems and making some philosophical remarks. The proof of Theorem 4 relies on a straightforward calculus of variations applied to the two variational problems in (1.14) and (1.17), and will be omitted. We refer the reader to [1], [4] for details.

2 Sketch of the derivation of Theorems 1, 2 and 3

2.a. Local times of random walk. Let

$$\begin{aligned} &S_0 = 0,\ S_n = X_1 + \ldots X_n,\ n \in \mathbb{N}, \\ &(X_i) \text{ i.i.d. with } P_h(X_i = 1) = 1 - P_h(X_i = 0) = h \end{aligned} \tag{2.1}$$

denote the random walk with drift h that serves as the underlying migration process of the particles. For $x \in \mathbb{N}_0$ and $n \in \mathbb{N}_0$, let

$$\ell_n(x) = |\{0 < i \leq n\colon\ S_i = x\}| \tag{2.2}$$

denote its local time at site x up to time n. The following representation of the local particle density in (1.9) is the starting point for our explanation. We write P_h and E_h to denote probability and expectation w.r.t. the random walk.

Lemma 1 $\hat{d}_n(F) = E_h\Big(\prod_{y\in\mathbb{N}_0} [b_{-y}]^{\ell_n(y)}\Big)$.

Proof. Fix F. From the evolution mechanism of our process (recall steps (i) and (ii) in Section 1.b) we obtain the recursion relation

$$E(\eta_n(x)|F) = (1-h)b_x E(\eta_{n-1}(x)|F) + hb_{x-1}E(\eta_{n-1}(x-1)|F). \tag{2.3}$$

In the notation of (2.1) this may be rewritten as

$$\hat{d}_n(x,F) = E_h\Big(b_{x-X_1}\hat{d}_{n-1}(x-X_1,F)\Big), \tag{2.4}$$

where $\hat{d}_n(x,F)$ is the local particle density at time n at site x. Iteration of (2.4) gives

$$\hat{d}_n(x,F) = E_h\Big(\prod_{i=1}^{n} b_{x-S_i}\hat{d}_0(x-S_n,F)\Big). \tag{2.5}$$

Finally, put $x=0$, substitute $\hat{d}_0(\cdot\,,F)\equiv 1$ and write $\prod_{i=1}^n b_{-S_i} = \prod_{y\in\mathbb{N}_0}[b_{-y}]^{\ell_n(y)}$. ∎

Next, we rewrite Lemma 1 in a form that is more appropriate for the large deviation analysis to be carried out in Section 2.b. For $x\in\mathbb{N}_0$, let

$$\ell(x) = \lim_{n\to\infty}\ell_n(x) = |\{0<i<\infty\colon\ S_i = x\}| \tag{2.6}$$

denote the total local time at site x.

Lemma 2 *For $h\in(0,1)$:*

$$\hat{d}_n(F) = \sum_{x\in\mathbb{N}_0} P_h(S_n = x)E_h(x,n,\tilde{F}), \tag{2.7}$$

where $\tilde{F}$ is the reversed environment defined by $\tilde{F}_y = F_{-y}$, $y\in\mathbb{Z}$, and

$$E_h(x,n,F) = E_h\Big(\prod_{y=0}^{x}[b_y]^{\ell(y)} \mid \sum_{y=0}^{x}\ell(y) = n\Big). \tag{2.8}$$

Proof. Rewrite the representation in Lemma 1 as in (2.7) with

$$E_h(x,n,F) = E_h\Big(\prod_{y=0}^{x}[b_y]^{\ell(y)} \mid S_n = x\Big), \tag{2.9}$$

using that $\ell_n(y)=0$ for all $y>x$ on the event $\{S_n = x\}$. Add $S_{n+1} = x+1$ to the condition in (2.9), note that $\{S_n = x, S_{n+1} = x+1\} = \{\sum_{y=0}^{x}\ell(y) = n\}$, and use that on this event $\ell_n(y) = \ell(y)$ for all $0\le y\le x$. ∎

2.b. Large deviations and growth rates. Lemma 2 is the starting point for our large deviation analysis.

Local growth rate: We begin by writing (2.7) as an integral:

$$\hat{d}_n(F) = n \int_{\theta\in(0,1]} d\theta\, P_h(S_n = \lceil\theta n\rceil) E_h(\lceil\theta n\rceil, n, \tilde{F}) + (1-h)^n [b_0]^n. \tag{2.10}$$

The last term is harmless because $\lambda \geq \log[M(1-h)]$ (recall (1.8), (1.11), (1.16) and (1.25)). Next, under the integral we may replace E_h by E_θ, i.e.,

$$E_h(\lceil\theta n\rceil, n, \tilde{F}) = E_\theta(\lceil\theta n\rceil, n, \tilde{F}). \tag{2.11}$$

Indeed, the condition $\{S_n = \lceil\theta n\rceil\}$ fixes the number of steps to the right up to time n at $\lceil\theta n\rceil$, which makes the conditional probability independent of the drift. Next, Sanov's Theorem in Section 1.a tells us that

$$\lim_{n\to\infty} \frac{1}{n} \log P_h(S_n = \lceil\theta n\rceil) = -H(B_\theta|B_h). \tag{2.12}$$

Indeed, pick $\mathbb{S} = \{0,1\}$, $\mu = B_h$, $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and note that $\{S_n = \lceil\theta n\rceil\} = \{L_n = B_{\lceil\theta n\rceil/n}\}$. Next, we will show that

$$\lim_{n\to\infty} \frac{1}{n} \log E_\theta(\lceil\theta n\rceil, n, \tilde{F}) = -\hat{J}_\beta(\theta) \qquad \mathbb{P}-a.s., \tag{2.13}$$

where $\hat{J}_\beta(\theta)$ is the inner variational problem in (1.17). Theorem 2 follows from (2.10), (2.12) and (2.13) with the help of Varadhan's Lemma (see [8], Theorem III.13).

To prove (2.13), we again use Sanov's Theorem. For $N \in \mathbb{N}$, let

$$L_N = \frac{1}{N}\sum_{x=0}^{N-1} \delta_{(\ell(x),b_x)} \tag{2.14}$$

denote the empirical distribution over the N-interval of the total local time process and the random environment jointly. In terms of this quantity, we may write

$$\begin{aligned}
\prod_{y=0}^{\lceil\theta n\rceil} [b_y]^{\ell(y)} &= \exp\Big[\sum_{y=0}^{\lceil\theta n\rceil} \ell(y)\, \log b_y\Big] \\
&= \exp\Big[K_n \sum_{i\in\mathbb{N}, j\in\mathbb{B}} L_{K_n}(i,j)\, i \log j\Big] \\
&= \exp[K_n \hat{f}(L_{K_n})]
\end{aligned} \tag{2.15}$$

with $\hat{f}$ defined in (1.18) and $K_n = \lceil\theta n\rceil + 1$. Let

$$\hat{\mathcal{A}}_n = \Big\{\nu \in \mathcal{P}(\mathbb{N}\times\mathbb{B})\colon \sum_{i\in\mathbb{N}, j\in\mathbb{B}} i\nu(i,j) = \frac{n}{K_n}\Big\}. \tag{2.16}$$

Since $\{\sum_{y=0}^{\lceil\theta n\rceil} \ell(y)\} = \{L_{K_n} \in \hat{\mathcal{A}}_n\}$, we get from (2.8), (2.11) and (2.15) that

$$E_\theta(\lceil\theta n\rceil, n, \tilde{F}) = \int e^{K_n \hat{f}(\nu)} P_\theta(L_{K_n} \in d\nu \mid L_{K_n} \in \hat{\mathcal{A}}_n). \tag{2.17}$$

Now, under P_θ the sequence $(\ell(x), b_x)_{x\in\mathbb{N}_0}$ is i.i.d. with state space $\mathbb{S} = \mathbb{N}\times\mathbb{B}$ and marginal law $\mu = \pi_\theta \times \beta$. Therefore $(L_N)_{N\in\mathbb{N}}$ satisfies the large deviation principle on $\mathcal{P}(\mathbb{N}\times\mathbb{B})$ with rate function $H(\nu|\pi_\theta\times\beta)$. Hence (2.13) follows from (2.17) via Varadhan's Lemma.

There are two technical points here that need consideration. First, we are in fact using Sanov's Theorem under two conditions, namely,

$$\begin{aligned}
\sum_{i\in\mathbb{N},j\in\mathbb{B}} iL_{K_n}(i,j) &= \tfrac{n}{K_n}, \\
\sum_{i\in\mathbb{N}} L_{K_n}(i,j) &= \tfrac{1}{K_n}\sum_{x=0}^{K_n-1}\delta_{b_x}(j) \qquad \forall j\in\mathbb{B},
\end{aligned} \tag{2.18}$$

with the last empirical measure fixed because F is fixed. However, as $n\to\infty$ the right-hand sides of (2.18) tend to $1/\theta$ resp. $\beta(j)$ (by the ergodic theorem applied to F). The fact that this translates into the two restrictions in $\hat{\mathcal{M}}_{\theta,\beta}$ defined in (1.18) requires some continuity arguments. Second, the application of Varadhan's Lemma requires certain regularity properties, which can be obtained via approximation. We refer the reader to [1], [4] for details.

Global growth rate: Since, by (1.9),

$$d_n = \mathbb{E}(\hat{d}_n(F)), \tag{2.19}$$

Lemma 1 also gives us a representation of the global particle density. We can therefore follow the same line of argument as above. Because of (2.19), we put

$$E_h(\lceil\theta n\rceil, n) = \mathbb{E}\Big(E_h(\lceil\theta n\rceil, n, \tilde{F})\Big). \tag{2.20}$$

It suffices to show that the analogue of (2.13) holds:

$$\lim_{n\to\infty}\frac{1}{n}\log E_\theta(\lceil\theta n\rceil, n) = -J_\beta(\theta) \tag{2.21}$$

with $J_\beta(\theta)$ the inner variational problem in (1.14). For $N\in\mathbb{N}$, let

$$L_N = \frac{1}{N}\sum_{x=0}^{N-1}\delta_{\ell(x)}. \tag{2.22}$$

We have

$$\begin{aligned}
\mathbb{E}\Big(E_h(\lceil\theta n\rceil, n, \tilde{F})\Big) &= \mathbb{E}\Big(E_\theta(\lceil\theta n\rceil, n, \tilde{F})\Big) \\
&= E_\theta\Big(\mathbb{E}\Big(\prod_{y=0}^{\lceil\theta n\rceil}[b_{-y}]^{\ell(y)}\Big) \mid \sum_{y=0}^{\lceil\theta n\rceil}\ell(y) = n\Big) \\
&= E_\theta\Big(\prod_{y=0}^{\lceil\theta n\rceil}\sum_{j\in\mathbb{B}}\beta(j)j^{\ell(y)} \mid \sum_{y=0}^{\lceil\theta n\rceil}\ell(y) = n\Big) \\
&= \int e^{K_n f(L_{K_n})} P_\theta(L_{K_n}\in d\nu \mid L_n\in\mathcal{A}_n)
\end{aligned} \tag{2.23}$$

with f defined in (1.15), $K_n = \lceil\theta n\rceil + 1$ and

$$\mathcal{A}_n = \Big\{\nu\in\mathcal{P}(\mathbb{N})\colon \sum_{i\in\mathbb{N}} i\nu(i) = \frac{n}{K_n}\Big\}. \tag{2.24}$$

Under P_θ the sequence $(\ell(x))_{x\in\mathbb{N}_0}$ is i.i.d. with state space $\mathbb{S}=\mathbb{N}$ and marginal law $\mu=\pi_\theta$. Therefore $(L_N)_{N\in\mathbb{N}}$ satisfies the large deviation principle on $\mathcal{P}(\mathbb{N})$ with rate function $H(\nu|\pi_\theta)$. Hence (2.21) follows from (2.20) and (2.23) via Varadhan's Lemma.

This completes the sketch of the proof of Theorems 1 and 2.

2.c. Relation between the global and the local growth rate. The proof of Theorem 3 runs as follows. For any $\nu \in \hat{\mathcal{M}}_{\theta,\gamma}$, we have the identity

$$H(\nu|\pi_\theta \times \gamma) + H(\gamma|\beta) = H(\nu|\pi_\theta \times \beta) \tag{2.25}$$

because $\sum_{i\in\mathbb{N}} \nu(i,j) = \gamma(j)\ \forall j \in \mathbb{B}$. Hence the right-hand side of (1.19) equals

$$\sup_{\gamma\in\mathcal{P}(\mathbb{B})} \sup_{\nu\in\hat{\mathcal{M}}_{\theta,\gamma}} [\hat{f}(\nu) - H(\nu|\pi_\theta \times \beta)]. \tag{2.26}$$

Let $\bar{\nu}(i) = \sum_{j\in\mathbb{B}} \nu(i,j)$. A little computation gives

$$[\hat{f}(\nu) - H(\nu|\pi_\theta \times \beta)] = [f(\bar{\nu}) - H(\bar{\nu}|\pi_\theta)] - D \tag{2.27}$$

with

$$D = \sum_{i\in\mathbb{N}} \bar{\nu}(i) H(\mu_i|\mu_i^*), \tag{2.28}$$

where $\mu_i(j) = \nu(i,j)/\bar{\nu}(i)$ and $\mu_i^*(j) = \beta(j)j^i / \sum_{j\in\mathbb{B}} \beta(j)j^i$. Since

$$\overline{\Big\{ \bigcup_{\gamma\in\mathcal{P}(\mathbb{B})} \hat{\mathcal{M}}_{\theta,\gamma} \Big\}} = \mathcal{M}_\theta, \tag{2.29}$$

there is no constraint on (μ_i), only on $\bar{\nu}$. Hence the supremum in (2.26) after we substitute (2.27) is taken at $(\mu_i) = (\mu_i^*)$ where $D = 0$. Therefore the right-hand side of (1.19) equals

$$\sup_{\bar{\nu}\in\mathcal{M}_\theta} [f(\bar{\nu}) - H(\bar{\nu}|\pi_\theta)], \tag{2.30}$$

which is the same as the left-hand side of (1.19).

3 Two extensions

(1) In [5] the same model as in Sections 1 and 2 is considered, but with a different initial state, namely,

$$\begin{aligned} \eta_0(x) &= 1 \quad \text{for } x \le 0, \\ &= 0 \quad \text{for } x > 0. \end{aligned} \tag{3.1}$$

In this situation we get *wave front propagation*. The exponential growth rate at a site moving at speed $\tau \ge 0$ is computed, again in terms of a variational problem arising via Sanov's Theorem. The result reveals two characteristic wave front speeds: τ_1, the speed of the front of *zero growth* (= the speed of the right-most particle) and τ_2, the speed of the front of *maximal growth* (= the speed up to which the growth rate equals $\hat{\lambda}(\beta, h)$). The latter speed exhibits a phase transition, changing from zero to positive as the drift in the migration crosses a threshold.

(2) In [2], [6], [7] the same model as in Sections 1 and 2 is considered, but with the migration running both forward and backward: particles jump one lattice spacing to the right with probability $\frac{1}{2}(1+h)$ or one lattice spacing to the left with probability $\frac{1}{2}(1-h)$, where $h \in (0,1)$

is again the drift parameter. The analogues of Theorems 1, 2, 3 and 4 are proved, but the analysis is considerably more complex than in the one-sided case. The reason is that the local time process no longer has an i.i.d. structure. The global and the local growth rate come out as variational problems, but this time a more refined version of Sanov's Theorem is needed, where the empirical measure in (1.1) is replaced by the so-called empirical process, i.e., the empirical measure built on a sequence of random variables rather than single random variables (see [3], Section 6). It turns out that the variational problems can be solved in terms of certain random continued fractions. Though the latter may seem a bit prohibitive, they can actually be used to derive a phase diagram similar to Figs. 1 and 2.

4 Conclusion

The model described in this paper is an example of an interacting particle system in a random medium. It is closely related to the class of problems studied in Sznitman's book [9], which revolves around large deviation theory for Brownian motion among Poissonian obstacles. The distinction between local and global corresponds to what in the physics literature is often called quenched and annealed, namely, in a fixed medium resp. an averaged medium.

An interesting open problem is to prove the law of large numbers

$$\frac{\eta_n(0)}{E(\eta_n(0)|F)} \to 1 \qquad (\mathbb{P} \times P) - a.s. \text{ as } n \to \infty. \tag{4.1}$$

Indeed, only then does $\hat{\lambda}(F)$ in (1.11) truly deserve the name of local growth rate, like λ in (1.11) deserves the name of global growth rate because of (1.10). No doubt (4.1) holds true because the particles at the origin at time n come from many different ancestors at time 0, but the proof has not been worked out.

One of the major challenges is to extend the results described above to higher dimension: particles migrate and branch on $\mathbb{Z}^d$ with $d \geq 2$. Qualitatively, we expect a phase diagram similar to Figs. 1 and 2, but it is not a priori clear that all the phase transitions survive in higher dimension. In any case, the variational study has not been carried through in any detail.

Large deviation theory is the crossroad of probability theory, ergodic theory and statistical physics. The sketch in Section 2 shows that both the global and the local growth rate are the result of an optimal selection from possible growth strategies (parametrized by θ and ν) for the path of descent of a typical particle (i.e., a particle drawn randomly from the global resp. the local population). *Relative entropy is the key notion, since it determines the weight of the different growth strategies and leads to the variational problems embodying the selection.*

As is evident from some of the other contributions in this volume, relative entropy shows up in various different contexts, but each time its role is one of selection. The model described in the present paper is but one illustration of this role. The rich behavior seen in Figs. 1 and 2 shows that relative entropy is capable of catching subtle phenomena like phase transitions, which makes it a versatile tool indeed.

References

[1] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, A variational approach to branching random walk in random environment, Ann. Probab. 21 (1993) 290–317.

[2] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On a variational problem for an infinite particle system in a random medium, Part I: The global growth rate, J. reine angew. Math. 454 (1994) 181–217.

[3] H.-O. Georgii, Aspects of entropy, in this volume.

[4] A. Greven and F. den Hollander, Population growth in random media, Part I: Variational formula and phase diagram, J. Stat. Phys. 65 (1991) 1123–1146.

[5] A. Greven and F. den Hollander, Population growth in random media, Part II: Wave front propagation, J. Stat. Phys. 65 (1991) 1147–1154.

[6] A. Greven and F. den Hollander, Branching random walk in random environment: phase transitions for local and global growth rates, Probab. Theory Relat. Fields 91 (1992) 195–249.

[7] A. Greven and F. den Hollander, On a variational problem for an infinite particle system in a random medium, Part II: The local growth rate, Probab. Theory Relat. Fields 100 (1994) 301–328.

[8] F. den Hollander, *Large Deviations*, Fields Institute Monographs 14, American Mathematical Society, Providence RI, 2000.

[9] A.-S. Sznitman, *Brownian Motion, Obstacles and Random Media*, Springer, Berlin, 1998.