

A CONDITIONING PRINCIPLE FOR GALTON–WATSON TREES

NATHANAËL BERESTYCKI, PETER MÖRTERS AND NADIA SIDOROVA

Abstract: We show that an infinite Galton-Watson tree, conditioned on its martingale limit being smaller than ε , converges as $\varepsilon \downarrow 0$ in law to the regular μ -ary tree, where μ is the essential minimum of the offspring distribution. This gives an example of entropic repulsion where the limit has no entropy.

1. INTRODUCTION AND STATEMENT OF THE RESULT

The problem of *conditioning principles* can be formulated in the following way: Given that some quantity averaged over a large number of individual random variables shows highly unlikely behaviour, describe the conditional law of an individual sample. This situation arises frequently in statistical mechanics, where the random variables describe individual features of particles (for example their velocity) and the ensemble of particles is subject to some constraint (for example a fixed energy per particle). The distribution of the individual feature given the constraint is then referred to as the *micro-canonical* distribution of the system. The most famous result in this respect is the *Gibbs conditioning principle*, which loosely speaking says that under the condition that the empirical measure

$$L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

of a family of independent random variables X_1, \dots, X_n with law P belongs to some set A , the law of X_1 converges to the probability measure Q that minimizes the relative entropy $H(Q|P)$ subject to the constraint $Q \in A$. There exist several refinements of this result describing rigorously the precise asymptotic strategy by which the random variables realize the large deviation event $\{L_n \in A\}$. See the book of Dembo and Zeitouni [10] for more on the classical Gibbs conditioning principle, [9, 11, 17] for refinements, and [12, 14, 15] for further examples of conditioning principles.

The conditioning principle of the present paper deals with Galton–Watson trees with a nondegenerate offspring variable N satisfying $P(N = 0) = 0$ and $EN \log N < \infty$. Let $a := EN$ be the mean offspring number. We denote by $(Z_n : n = 0, 1, \dots)$ the sequence of generation sizes of the Galton–Watson tree and note that by definition $Z_0 = 1$. By the Kesten-Stigum theorem the *martingale limit*

$$W := \lim_{n \rightarrow \infty} \frac{Z_n}{a^n}$$

is well-defined and strictly positive almost surely. Note that W can be seen as a random constant factor in front of a deterministic exponential growth term a^n , which together determine the asymptotics of the generation size Z_n . We are interested in the limit behaviour of Z_1 , or more generally of the entire tree, when we condition on the large deviation event that the martingale limit W is smaller than some $\varepsilon \downarrow 0$.

For the formulation of the result we denote by \mathcal{T} the space of all rooted trees with the property that every vertex has finite degree. A metric d on this space is uniquely determined by the requirement that $d(T_1, T_2) = e^{-n}$, when n is maximal with the property that the trees T_1 and T_2 coincide up to the n th generation. This makes (\mathcal{T}, d) a complete, separable metric space.

Theorem 1. *Suppose N is a random variable on the positive integers satisfying the condition $EN \log N < \infty$, and denote*

$$\mu := \min \{n \in \mathbb{N}: \mathbb{P}(N = n) > 0\} \geq 1.$$

Suppose that T is a Galton–Watson tree with offspring variable N and that W is the associated martingale limit. Then, as $\varepsilon \downarrow 0$, conditionally on the event $\{W < \varepsilon\}$ the tree T converges in law on (\mathcal{T}, d) to the regular μ -ary tree. Equivalently, for all $k \in \mathbb{N}$,

$$\lim_{\varepsilon \downarrow 0} \mathbb{P}(Z_k = \mu^k \mid W < \varepsilon) = 1,$$

where Z_k denotes the size of the k th generation.

From the point of view of large deviations theory this result is quite surprising, at least at a first glance. One would expect that the limiting behaviour represents the optimal strategy by which the event $W = 0$ is realized and that this strategy depends on the details of the law of N . Moreover, there seems to be no good reason why in the limit the *growth rate* of the tree should drop dramatically, or in fact why it should drop at all, as we only require the constant to be small. Above all, the probability of seeing a μ -ary tree up to the n th generation may be arbitrarily small and can certainly be much smaller than those of other trees with $Z_n \leq \varepsilon a^n$.

This becomes even more intriguing if the result is put in the context of *entropic repulsion*. This is an expression used by physicists to convey the idea that entropy maximisation may force certain systems to obey properties that are not obviously imposed on them *a priori*. This phenomenon has been studied mathematically in the context of the two-dimensional harmonic crystal with hard wall repulsion by Bolthausen et al [7], where the following result was proved. Consider the discrete Gaussian free field $(\varphi_x)_{x \in D_n}$ on a planar domain with mesh size $1/n$. If the field is conditioned to be nonnegative everywhere, then the typical value of the field φ_x at any point x in the interior of the domain will be highly concentrated near the value $(4/\pi) \log n$ with overwhelming probability as $n \rightarrow \infty$, and in particular under this conditioning the value of φ_x diverges to infinity. An analogous phenomenon is studied by Benjamini and Berestycki [2] and [3], where it is shown that conditioning a one-dimensional Brownian motion on some self-repelling behaviour may force the process to satisfy a strongly amplified version of the constraint. Usually, the reason entropic repulsion may arise is in order to increase the entropy of the system, i.e., make room for fluctuations. Thus the eventual state of the system is a compromise between the energy cost of adopting an unusual behaviour and the entropic benefits. Theorem 1 may also be cast in this framework, as it shows that the effect of requiring the constant W to be small is to change the overall exponential growth rate from a to μ . However, if the limiting state of the system is the regular μ -ary tree, which is non-random, what could the entropic benefits possibly be?

The resolution of this apparent paradox comes from understanding the inhomogeneity of the optimal strategy, and can be explained by a closer look at the formula

$$Z_n \sim W a^n.$$

While the growth rate $\log a$ is purely asymptotic, i.e. depends only on the offspring numbers *after* any given generation, the growth constant W depends heavily on the *initial* generations of the tree. It turns out that, roughly speaking, the collection of trees which form the optimal strategy to achieve $W < \varepsilon$ have minimal offspring for a few generations, the exact number depending on ε , and causes high entropic and energetic cost but only for a small number of generations, and then switch to growth with the natural rate $\log a$. The initial behaviour ensures that W is small at a minimal probabilistic cost, because for all but a finite number of generations the trees can have their natural growth. The topology on \mathcal{T} compares trees starting from their root so that in the limit we only see the behaviour in the initial generations. This leads to a limiting object with minimal growth rate at all generations and creates the illusion of a drop in the growth rate for the optimal strategy. A somewhat similar phenomenon is observed by Bansaye and Berestycki [1] in the context of branching processes in random environment, although they consider situations where the growth rate is directly conditioned to be atypical.

Our interest in this ‘paradox’ does not come from the study of trees alone. Indeed, Mörters and Ortgiese [16] describe a range of problems, mostly related to local times of Brownian motion, which have a similar intrinsic structure and could therefore also satisfy loosely analogous conditioning principles. However, we shall defer the discussion of such problems to a different place as, unlike in these problems, the main mathematical difficulty here is related to the discrete nature of the distribution N .

2. PROOF OF THEOREM 1

Denote $p_k := P(N = k)$ and recall that by our definitions $p_0 = \dots = p_{\mu-1} = 0$, $p_\mu > 0$. The basic idea of the proof is to combine tail asymptotics at zero for the random variable W with the self-similarity property of Galton–Watson trees, which states that, for every $n \in \mathbb{N}$,

$$W = \frac{1}{a^n} \sum_{i=1}^{Z_n} W_i, \quad (1)$$

where W_i , $i = 1, 2, \dots$ are independent variables with the same law as W , independent of Z_n . This follows easily from the decomposition of the tree according to the ancestry in the n th generation.

We first give the proof in the case $\mu = 1$, which is very simple as in this case the μ -ary tree is degenerated and has no exponential growth. In this case the tail at zero of the random variable W is fat, more precisely there exist constants $0 < c < C$ such that

$$c \varepsilon^\tau \leq \mathbb{P}(W < \varepsilon) \leq C \varepsilon^\tau \quad \text{for all } 0 < \varepsilon < 1,$$

where $\tau := -\log p_1 / \log a$, see [16, Theorem 1(a)] for a simple proof. Using (1) we infer

$$\begin{aligned} \mathbb{P}(Z_n > 1, W < \varepsilon) &\leq P(W_1 + W_2 < a^n \varepsilon) \leq P(W < a^n \varepsilon)^2 \\ &\leq C^2 (a^n \varepsilon)^{2\tau} \leq \mathbb{P}(W < \varepsilon) \left(\frac{C^2}{c} a^{2n\tau}\right) \varepsilon^\tau, \end{aligned}$$

and hence

$$\mathbb{P}(Z_n > 1 \mid W < \varepsilon) \leq \left(\frac{C^2}{c} a^{2n\tau}\right) \varepsilon^\tau \xrightarrow{\varepsilon \downarrow 0} 0,$$

as required to complete the proof in the case $\mu = 1$.

Now we consider the case $\mu > 1$ and assume that $p_\mu \neq 1$ to avoid trivialities. We define the *Böttcher constant* $\beta \in (0, 1)$ by

$$a^\beta = \mu.$$

A function $V: (0, \infty) \rightarrow (0, \infty)$ is called *multiplicatively periodic* with period $\lambda \neq 1$ if $V(\lambda x) = V(x)$ for all $x > 0$. Biggins and Bingham [4, Theorem 3] show that there exists a real-analytic multiplicatively periodic function $M: (0, \infty) \rightarrow (0, \infty)$ with period $a^{1-\beta} = a/\mu > 1$ such that

$$-\log \mathbb{P}(W < x) = M(x) x^{-\frac{\beta}{1-\beta}} + o(x^{-\frac{\beta}{1-\beta}}) \quad \text{as } x \downarrow 0, \quad (2)$$

see also Fleischmann and Wachtel [13] for refinements of this statement. A key argument in the proof of (2) is to relate the left tail of a positive random variable to the behaviour of its Laplace transform at infinity in a way reminiscent of the Tauberian theorem of de Bruijn, see [6, Theorem 4.12.9]. In the next lemma we generalise this result to the case of several independent copies of W using the same basic method of proof as in [4].

Lemma 2. *Let X be a positive random variable such that, for some $a > 1$ and $\beta \in (0, 1)$, for all $s > 0$,*

$$\lim_{n \rightarrow \infty} \frac{\log \mathbb{E} \exp\{-sa^n X\}}{a^{n\beta}} =: k(s)$$

with some real-analytic function k on $(0, \infty)$. Then there exists a real-analytic multiplicatively periodic function $V: (0, \infty) \rightarrow (0, \infty)$ with period $a^{1-\beta}$ such that, for any $m \in \mathbb{N}$ and X_1, \dots, X_m independent with the same distribution as X , we have

$$-\log \mathbb{P}(X_1 + \dots + X_m < x) = m V(x/m)(x/m)^{-\frac{\beta}{1-\beta}} + o(x^{-\frac{\beta}{1-\beta}}) \quad \text{as } x \downarrow 0.$$

Proof. Let Y_n be real-valued random variables and denote, for $s > 0$,

$$k_n(s) := \log \mathbb{E} \exp\{sY_n\} \in (-\infty, \infty].$$

Assume that for some sequence of positive numbers $b_n \uparrow \infty$, we have

$$\lim_{n \rightarrow \infty} \frac{k_n(s)}{b_n} =: \hat{k}(s) \in (-\infty, \infty].$$

The Fenchel dual of \hat{k} is given by $\hat{k}^*(x) := \sup_{s>0} \{xs - \hat{k}(s)\} \in (-\infty, \infty]$. By a variant of the Gärtner–Ellis theorem, see [4, Corollary 1], we have

$$\lim_{n \rightarrow \infty} \frac{-\log \mathbb{P}(Y_n \geq b_n y)}{b_n} = \hat{k}^*(y) \quad \text{for all } y \in (\lim_{s \downarrow 0} \hat{k}'(s), \lim_{s \uparrow \infty} \hat{k}'(s)).$$

We apply this first to the sequence $Y_n := -a^n X$ and observe that $\hat{k} = k$ in this case. Note that k satisfies, by definition, $k(s) = v(s)s^\beta$ for a multiplicatively periodic function v with period a . Using that k is strictly convex and decreasing we get $\lim_{s \downarrow 0} k'(s) =: -\delta < 0$, and

$$0 \geq \lim_{s \uparrow \infty} k'(s) = \lim_{n \rightarrow \infty} k'(a^{n+1}) \geq \lim_{n \rightarrow \infty} \frac{k(a^{n+1}) - k(a^n)}{a^{n+1} - a^n} = \lim_{n \rightarrow \infty} a^{n(\beta-1)} v(1) \frac{a^\beta - 1}{a - 1} = 0.$$

Therefore,

$$\lim_{n \rightarrow \infty} \frac{-\log \mathbb{P}(-a^n X \geq a^{n\beta} y)}{a^{n\beta}} = k^*(y) \quad \text{for all } y \in (-\delta, 0),$$

Setting $x = -y$ and rearranging,

$$\lim_{n \rightarrow \infty} \frac{-\log \mathbb{P}(X \leq a^{-n(1-\beta)} x)}{a^{n\beta} x^{-\beta/(1-\beta)}} = k^*(-x) x^{\frac{\beta}{1-\beta}} =: V(x), \quad \text{for } x \in (0, \delta),$$

where V is real-analytic and multiplicatively periodic with period $a^{1-\beta}$.

Now, consider $\tilde{Y}_n = -a^n(X_1 + \cdots + X_m)$. We have

$$\lim_{n \rightarrow \infty} \frac{\log \mathbb{E} \exp\{-sa^n(X_1 + \cdots + X_m)\}}{a^{n\beta}} = m \lim_{n \rightarrow \infty} \frac{\log \mathbb{E} \exp\{-sa^n X\}}{a^{n\beta}} = mk(s),$$

and hence

$$\lim_{n \rightarrow \infty} \frac{-\log \mathbb{P}(-a^n(X_1 + \cdots + X_m) \geq a^{n\beta}y)}{a^{n\beta}} = \sup_{s>0} \{ys - mk(s)\} = mk^*(y/m)$$

for all $y \in (-m\delta, 0)$. Setting $x = -y$ and rearranging we obtain, for all $x \in (0, m\delta)$,

$$\lim_{n \rightarrow \infty} \frac{-\log \mathbb{P}(X_1 + \cdots + X_m \leq a^{-n(1-\beta)}x)}{a^{n\beta}x^{-\beta/(1-\beta)}} = mk^*(-x/m)x^{\frac{\beta}{1-\beta}} = V(x/m)m^{\frac{1}{1-\beta}}.$$

Denote by $H_n(x)$ the fraction on the left hand side, and by $H(x)$ the right hand side of the display above. Further, denote

$$\hat{H}_n(x) := H_n(x)x^{-\frac{\beta}{1-\beta}} \quad \text{and} \quad \hat{H}(x) := H(x)x^{-\frac{\beta}{1-\beta}}.$$

Let $I = [a^{-2(1-\beta)}m\delta, a^{-(1-\beta)}m\delta]$. Then \hat{H}_n converges to \hat{H} pointwise on I , \hat{H} is continuous on I , and each \hat{H}_n is decreasing. Hence \hat{H}_n converges to \hat{H} uniformly on I , see e.g. [8], and therefore H_n converges uniformly to H on I . By the periodicity of V we have

$$\sup_{x \leq a^{-n(1-\beta)}I} \left| x^{\frac{\beta}{1-\beta}} \log \mathbb{P}(X_1 + \cdots + X_m < x) + V(x/m)m^{\frac{1}{1-\beta}} \right| = \sup_{x \in I} |H_n(x) - H(x)|,$$

and hence

$$\begin{aligned} & \sup_{x \leq a^{-N(1-\beta)}m\delta} \left| x^{\frac{\beta}{1-\beta}} \log P(X_1 + \cdots + X_m < x) + V(x/m)m^{\frac{1}{1-\beta}} \right| \\ &= \sup_{n>N} \sup_{x \in I} |H_n(x) - H(x)|, \end{aligned}$$

which converges, as $N \rightarrow \infty$, to zero as required. \square

The next lemma states a basic property of real analytic, multiplicatively periodic functions.

Lemma 3. *Let $V: (0, \infty) \rightarrow (0, \infty)$ be a real-analytic, multiplicatively periodic function and let $\gamma > 0$. Suppose that B be a dense subset of $[1, \infty)$ such that*

$$\liminf_{\varepsilon \rightarrow 0} (V(\varepsilon/b)b^\gamma - V(\varepsilon)) \geq 0 \quad \text{for all } b \in B.$$

Then, for any $b_0 > 1$,

$$\inf_{\substack{\varepsilon > 0 \\ b \geq b_0}} (V(\varepsilon/b)b^\gamma - V(\varepsilon)) > 0.$$

Proof. Define a real-analytic function $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(y) = \log V(e^{-y})$. Then

$$V(\varepsilon/b)b^\gamma - V(\varepsilon) = \exp\{f(-\log \varepsilon + \log b)\}b^\gamma - \exp\{f(-\log \varepsilon)\},$$

and, substituting $x = -\log \varepsilon$ and $\delta = \log b$, we obtain that it suffices to show that

$$\liminf_{x \rightarrow \infty} (e^{f(x+\delta)+\gamma\delta} - e^{f(x)}) \geq 0 \quad \text{for all } \delta \in D := \{\log b: b \in B\} \quad (3)$$

$$\implies \inf_{\substack{x \in \mathbb{R} \\ \delta \geq \delta_0}} (e^{f(x+\delta)+\gamma\delta} - e^{f(x)}) > 0 \quad \text{for any } \delta_0 > 0. \quad (4)$$

By periodicity of f , the statement of (3) is equivalent to $f(x + \delta) + \gamma\delta \geq f(x)$ for all $x \in \mathbb{R}$ and $\delta \in D$. As B is dense in $[1, \infty)$, D is dense in $[0, \infty)$ and, using the continuity of f ,

$$f(x + \delta) + \gamma\delta \geq f(x) \quad \text{for all } x \in \mathbb{R} \text{ and } \delta \geq 0. \quad (5)$$

Suppose that (4) is not true and the infimum is equal to zero. Since f is periodic and

$$\liminf_{\delta \uparrow \infty} \inf_{x \in \mathbb{R}} \{e^{f(x+\delta)+\gamma\delta} - e^{f(x)}\} = \infty,$$

the infimum in (4) is attained at some point $(\hat{x}, \hat{\delta})$. As this infimum is zero, we infer that

$$f(\hat{x} + \hat{\delta}) + \gamma\hat{\delta} = f(\hat{x}). \quad (6)$$

Let $\eta \in [0, \hat{\delta}]$. Using (5) for the points \hat{x}, η and $\hat{x} + \eta, \hat{\delta} - \eta$, we obtain $f(\hat{x} + \eta) + \gamma\eta \geq f(\hat{x})$ and $f(\hat{x} + \hat{\delta}) + \gamma(\hat{\delta} - \eta) \geq f(\hat{x} + \eta)$. The second inequality together with (6) implies $f(\hat{x}) \geq f(\hat{x} + \eta) + \gamma\eta$, which together with the first inequality gives $f(\hat{x} + \eta) = f(\hat{x}) - \gamma\eta$ for all $\eta \in [0, \hat{\delta}]$. Hence f is linear and non-zero on $[\hat{x}, \hat{x} + \hat{\delta}]$. As it is real-analytic it must be linear on \mathbb{R} , contradicting the periodicity of f . \square

We now return to the study of the martingale limit W . A result from [5] states that

$$\lim_{n \rightarrow \infty} \frac{\log \mathbb{E} \exp\{-sa^n W\}}{a^{n\beta}} = k(s)$$

for some real-analytic function k on $(0, \infty)$. Using Lemma 2 we infer from this that, for some real-analytic and multiplicatively periodic $M: (0, \infty) \rightarrow (0, \infty)$ with period $a^{1-\beta}$ we have, for any $m \in \mathbb{N}$,

$$-\log \mathbb{P}(W_1 + \dots + W_m < \varepsilon) = m M(\varepsilon/m) (\varepsilon/m)^{-\frac{\beta}{1-\beta}} + o(\varepsilon^{-\frac{\beta}{1-\beta}}) \quad \text{as } \varepsilon \downarrow 0. \quad (7)$$

Using first (1), then (7), and finally $a^\beta = \mu$ and periodicity of M , we get

$$\begin{aligned} \log \mathbb{P}(W < \varepsilon \mid Z_n = m) &= \log \mathbb{P}\left(\sum_{i=1}^m W_i < \varepsilon a^n\right) \\ &= -M(\varepsilon a^n/m) m^{\frac{1}{1-\beta}} a^{-\frac{\beta n}{1-\beta}} \varepsilon^{-\frac{\beta}{1-\beta}} + o(\varepsilon^{-\frac{\beta}{1-\beta}}) \\ &= -M(\varepsilon \mu^n/m) (m/\mu^n)^{\frac{1}{1-\beta}} \varepsilon^{-\frac{\beta}{1-\beta}} + o(\varepsilon^{-\frac{\beta}{1-\beta}}). \end{aligned}$$

Combining with (7) again we obtain

$$\begin{aligned} \log \mathbb{P}(W < \varepsilon) - \log \mathbb{P}(W < \varepsilon \mid Z_n = m) \\ = \left(M(\varepsilon \mu^n/m) (m/\mu^n)^{\frac{1}{1-\beta}} - M(\varepsilon)\right) \varepsilon^{-\frac{\beta}{1-\beta}} + o(\varepsilon^{-\frac{\beta}{1-\beta}}). \end{aligned} \quad (8)$$

Lemma 3 enables us to analyse the bracketed term.

Lemma 4. *For any $b_0 > 1$, we have*

$$\inf_{\substack{\varepsilon > 0 \\ b \geq b_0}} (M(\varepsilon/b) b^{\frac{1}{1-\beta}} - M(\varepsilon)) > 0.$$

Proof. Since M is real-analytic and multiplicatively periodic, it suffices to check that it satisfies the assumptions of Lemma 3 with $\gamma = 1/(1-\beta)$. For fixed $n \in \mathbb{N}$ we define

$$B_n := \{m/\mu^n : \mathbb{P}(Z_n = m) \neq 0\}$$

and $B = \cup_{n \in \mathbb{N}} B_n$. We now show that B is dense in $[1, \infty)$. Indeed, as $p_\mu \neq 1$ there exists $\nu > \mu$ with $p_\nu \neq 0$. Denote $d = \nu - \mu$. We can prove by induction that

$$A_n := \{m/\mu^n : \mu^n \leq m \leq \nu^n, m \equiv \mu^n \pmod{d}\} \subset B_n.$$

This is obvious for $n = 1$. Assuming that $\mathbb{P}(Z_{n-1} = r) \neq 0$ for all r such that $\mu^{n-1} \leq r \leq \nu^{n-1}$, $r \equiv \mu^{n-1} \pmod{d}$ we obtain that $\mathbb{P}(Z_n = m) \neq 0$ for all m such that there is r satisfying the conditions above and such that $\mu r \leq m \leq \nu r$, $m \equiv \mu^n \pmod{d}$. It is easy to

see that this is equivalent to the condition $\mu^n \leq m \leq \nu^n, m \equiv \mu^n \pmod{d}$. Hence $A_n \subset B_n$ and since $\cup_{n \in \mathbb{N}} A_n$ is dense in $[1, \infty)$ we obtain that $\cup_{n \in \mathbb{N}} B_n$ is dense in $[1, \infty)$.

Let $b \in B$, that is, $b = m/\mu^n$ for some m and n . We have

$$\mathbb{P}(Z_n = m | W < \varepsilon) \mathbb{P}(W < \varepsilon) = \mathbb{P}(W < \varepsilon, Z_n = m) = \mathbb{P}(W < \varepsilon | Z_n = m) \mathbb{P}(Z_n = m)$$

and so

$$\liminf_{\varepsilon \downarrow 0} \frac{\mathbb{P}(W < \varepsilon)}{\mathbb{P}(W < \varepsilon | Z_n = m)} = \liminf_{\varepsilon \downarrow 0} \frac{\mathbb{P}(Z_n = m)}{\mathbb{P}(Z_n = m | W < \varepsilon)} \geq \mathbb{P}(Z_n = m) > 0.$$

Hence

$$\liminf_{\varepsilon \downarrow 0} \{ \log \mathbb{P}(W < \varepsilon) - \log \mathbb{P}(W < \varepsilon | Z_n = m) \} > -\infty. \quad (9)$$

Combining (8) with (9) we obtain

$$\liminf_{\varepsilon \downarrow 0} \left(M(\varepsilon/b) b^{\frac{1}{1-\beta}} - M(\varepsilon) \right) = \liminf_{\varepsilon \downarrow 0} \left(M(\varepsilon \mu^n / m) (m/\mu^n)^{\frac{1}{1-\beta}} - M(\varepsilon) \right) \geq 0,$$

as required. \square

We now complete the proof of Theorem 1. Fix $n \in \mathbb{N}$ and use Lemma 4 to find $c > 0$ (depending on n) such that $M(\varepsilon/b) b^{\frac{1}{1-\beta}} - M(\varepsilon) \geq 2c$ for all $\varepsilon > 0$ and $b \geq 1 + \mu^{-n}$. For all $m \geq \mu^n + 1$ we have $m/\mu^n \geq 1 + \mu^{-n}$ and hence (8) implies

$$\log \mathbb{P}(W < \varepsilon) - \log \mathbb{P}(W < \varepsilon | Z_n = m) \geq c\varepsilon^{-\frac{\beta}{1-\beta}}.$$

Therefore

$$\begin{aligned} \mathbb{P}(Z_n > \mu^n | W < \varepsilon) &= \sum_{m=\mu^n+1}^{\infty} \frac{\mathbb{P}(W < \varepsilon | Z_n = m)}{\mathbb{P}(W < \varepsilon)} \mathbb{P}(Z_n = m) \\ &\leq \sum_{m=\mu^n+1}^{\infty} \exp \left\{ -c\varepsilon^{-\frac{\beta}{1-\beta}} \right\} \mathbb{P}(Z_n = m) \leq \exp \left\{ -c\varepsilon^{-\frac{\beta}{1-\beta}} \right\} \rightarrow 0 \end{aligned}$$

as $\varepsilon \downarrow 0$, completing the proof of Theorem 1 in the case $\mu > 1$.

Remark 1. Lemma 4 can be seen as an illustration of the near-constancy phenomenon (see [4] and references therein), which consists in the fact that the function M does not vary too much. Some numerical studies show that the variation of M can be very small, and close theoretical bounds for M are obtained for the case of an infinitely divisible distribution. No theoretical framework yet exists to describe the near-constancy of M in the general case. Lemma 4 implies that

$$M(x)(y/x)^{\frac{1}{1-\beta}} - M(y) \geq 0 \quad \text{for all } 0 < x \leq y,$$

and so, with $g(x) = x^{-\frac{1}{1-\beta}}$, the function $x \mapsto M(x)g(x)$ is decreasing. As the fluctuations of M do not destroy the monotonicity of the decreasing function g they cannot be too large.

Acknowledgement: The first author is supported by EPSRC grant EP/GO55068/1, and the second author is supported by an *Advanced Research Fellowship* from EPSRC.

REFERENCES

- [1] Bansaye, V., Berestycki, J.; Large deviations for branching processes in random environment. *Markov Proc. Rel. Fields.* 15, 493–524 (2009).
- [2] Benjamini, I., Berestycki, N.; Random paths with bounded local time. *J. Europ. Math. Soc.* to appear.
- [3] Benjamini, I., Berestycki, N.; An integral test for the transience of a Brownian path with limited local time. *Ann. Inst. H. Poinc. (B): Probab. Stat.*, to appear.
- [4] Biggins, J.D., Bingham, N.H.; Large deviations in the supercritical branching process. *Adv. Appl. Probab.* 25, 757–772 (1993).
- [5] Bingham, N.H.; On the limit of a supercritical branching process. *J. Appl. Probab.*, 25A, 215–228 (1988).
- [6] Bingham, N., Goldie, C., Teugels, J.; *Regular variation*. Cambridge University Press, Cambridge (1987).
- [7] Bolthausen, E., Deuschel, J.-D., Giacomin, G.; Entropic repulsion and the maximum of the two dimensional harmonic crystal. *Ann. Probab.*, 29, 1670–1692 (2001).
- [8] Buchanan, H.E., Hildebrandt, T.H.; Note on the convergence of a sequence of functions of a certain type. *Ann. Math. (2)*, 9, 123–126 (1908).
- [9] Csizsár, I.; Sanov property, generalized I -projection and a conditional limit theorem. *Ann. Probab.* 12, 768–793 (1984).
- [10] Dembo, A., Zeitouni, O.; *Large deviations techniques and applications*. Second edition. Applications of Mathematics 38. Springer-Verlag, New York, 1998.
- [11] Dembo, A., Zeitouni, O.; Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields* 104, 1–14 (1996).
- [12] Deuschel, J.-D., Stroock, D. W., Zessin, H.; Microcanonical distributions for lattice gases. *Comm. Math. Phys.* 139, 83–101 (1991).
- [13] Fleischmann, K., Wachtel, V.; On the left tail asymptotics for the limit law of supercritical Galton–Watson processes in the Böttcher case, *Ann. Inst. Henri Poincaré Probab. Stat.* 45, 201–225 (2009).
- [14] McKean, H. P., Vaninsky, K. L.; Brownian motion with restoring drift: the petit and micro-canonical ensembles. *Comm. Math. Phys.* 160, 615–630 (1994).
- [15] Meda, A., Ney, P.; The Gibbs conditioning principle for Markov chains. In: *Perplexing problems in probability*, 385–398, Progr. Probab., 44, Birkhäuser, Boston, 1999.
- [16] Mörters, P., Ortgiese, M.; Small value probabilities via the branching tree heuristic. *Bernoulli* 14, 277–299 (2008).
- [17] Stroock, D. W., Zeitouni, O.; Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In: *Random walks, Brownian motion, and interacting particle systems*, 399–424, Progr. Probab., 28, Birkhäuser, Boston, 1991.

Nathanaël Berestycki: Statistical Laboratory, DPMMS, University of Cambridge. Wilberforce Rd., Cambridge CB3 0WB. United Kingdom.

Peter Mörters: Department of Mathematical Sciences, University of Bath. Claverton Down, Bath BA2 7AY. United Kingdom.

Nadia Sidorova: Department of Mathematics, University College London. Gower Street, London WC1E 6BT. United Kingdom.