Bernoulli **27**(1), 2021, 615–636
https://doi.org/10.3150/20-BEJ1252

An invariance principle for biased voter model interfaces

RONGFENG SUN[1], JAN M. SWART[2] and JINJIONG YU[3]

[1]*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, 119076 Singapore. E-mail:* matsr@nus.edu.sg
[2]*The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod vodárenskou věží 4, 18200 Praha 8, Czech Republic. E-mail:* swart@utia.cas.cz
[3]*NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China. E-mail:* jinjiongyu@nyu.edu

We consider one-dimensional biased voter models, where 1's replace 0's at a faster rate than the other way round, started in a Heaviside initial state describing the interface between two infinite populations of 0's and 1's. In the limit of weak bias, for a diffusively rescaled process, we consider a measure-valued process describing the local fraction of type 1 sites as a function of time. Under a finite second moment condition on the rates, we show that in the diffusive scaling limit there is a drifted Brownian path with the property that all but a vanishingly small fraction of the sites on the left (resp. right) of this path are of type 0 (resp. 1). This extends known results for unbiased voter models. Our proofs depend crucially on recent results about interface tightness for biased voter models.

Keywords: biased voter model; branching and coalescing random walks; interface tightness; invariance principle

1. Introduction

1.1. Statement of the result

Let $\{0,1\}^{\mathbb{Z}}$ denote the space of all configurations of zeros and ones on $\mathbb{Z}$, that is, elements of $\{0,1\}^{\mathbb{Z}}$ are of the form $x = (x(i))_{i\in\mathbb{Z}}$ with $x(i) \in \{0,1\}$. The one-dimensional biased voter model $(X_t^\varepsilon)_{t\geq 0}$ with kernel $a(\cdot)$ and bias parameter $\varepsilon \in [0,1)$ is the interacting particle system with state space $\{0,1\}^{\mathbb{Z}}$ and formal generator

$$\begin{aligned} G^\varepsilon f(x) = {} & \sum_{i,j} a(j-i) 1_{\{x(i,j)=10\}} \big\{ f(x+e_j) - f(x) \big\} \\ & + (1-\varepsilon) \sum_{i,j} a(j-i) 1_{\{x(i,j)=01\}} \big\{ f(x-e_j) - f(x) \big\}, \end{aligned} \tag{1.1}$$

where $e_i(j) := 1_{\{i=j\}}$, and $x(i,j) = 10$ is shorthand for $x(i) = 1$, $x(j) = 0$. In words, (1.1) says that if $x(i) = 1$ and $x(j) = 0$, then the site j adopts the type of site i with rate $a(j-i)$. In the reverse case, when $x(i) = 0$ and $x(j) = 1$, the site j adopts the type of site i with rate $(1-\varepsilon)a(j-i)$. In particular, for $\varepsilon = 0$, we obtain a standard voter model.

The kernel a is a probability measure on $\mathbb{Z}$, not necessarily symmetric, such that $a(0) = 0$. In addition, throughout this paper, the following assumptions on a will always be in place:

1. a is irreducible, that is, each $k \in \mathbb{Z}$ can be written as a finite sum of $i \in \mathbb{Z}$ for which $a(i) > 0$,
2. a has mean zero, that is, $\sum_k a(k)k = 0$,
3. a has a finite second moment, that is, $\sigma^2 := \sum_k a(k)k^2 < \infty$.

1350-7265

We let

$$S^{01}_{\mathrm{int}} := \Big\{x \in \{0,1\}^{\mathbb{Z}} : \lim_{i\to-\infty} x(i) = 0,\ \lim_{i\to\infty} x(i) = 1\Big\} \tag{1.2}$$

denote the space of states in which an infinite population of 0's on the left and an infinite population of 1's on the right are separated by a hybrid zone containing a mixture of 0's and 1's. This hybrid zone is called the *interface* of the biased voter model. If X^ε is started from an initial state in S^{01}_{int} and a has a finite first moment, then it is known [4] that almost surely $X^\varepsilon_t \in S^{01}_{\mathrm{int}}$ for all $t \geq 0$.

If a has a finite second moment, then it is moreover known [13] that starting from a Heaviside state, the process returns to a Heaviside state after a random time with finite expectation, so that the process modulo translations is positive recurrent. This type of behavior is called *interface tightness* and was first proved for unbiased voter models in the groundbreaking paper [7] under a finite third moment condition on the rates. The authors of [5] improved this to a finite second moment condition, which they showed is optimal.

Let $\mathcal{M}(\mathbb{R})$ denote the space of locally finite measures on $\mathbb{R}$, equipped with the topology of vague convergence. We use $(X^\varepsilon_t)_{t\geq 0}$ to define a measure-valued process $(\mu^\varepsilon_t)_{t\geq 0}$ taking values in $\mathcal{M}(\mathbb{R})$ by

$$\mu^\varepsilon_t := \sum_{i\in\mathbb{Z}} \varepsilon X^\varepsilon_{\varepsilon^{-2}t}(i)\delta_{\varepsilon i} \quad (t \geq 0), \tag{1.3}$$

where δ_r denotes the delta-measure at $r \in \mathbb{R}$. We fix a standard Brownian motion $(W_t)_{t\geq 0}$ and define a Brownian motion $B = (B_t)_{t\geq 0}$ with drift $-\frac{1}{2}\sigma^2$ and diffusion coefficient σ^2 by $B_t := W_{\sigma^2 t} - \frac{1}{2}\sigma^2 t$. We use B to define a measure-valued process $(\mu_t)_{t\geq 0}$ by

$$\mu_t(\mathrm{d}x) := 1_{\{x\geq B_t\}}\,\mathrm{d}x \quad (t \geq 0), \tag{1.4}$$

that is, μ_t has the density $1_{[B_t,\infty)}$ with respect to Our main result says that $(\mu_t)_{t\geq 0}$ arises as the weak limit of $(\mu^\varepsilon_t)_{t\geq 0}$.

Theorem 1.1 (Invariance principle for biased voter model interface). *Fix $x \in S^{01}_{\mathrm{int}}$ and for $\varepsilon \in (0,1)$, let X^ε be the biased voter model with generator* (1.1) *and initial state x. Define $(\mu^\varepsilon_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ as in* (1.3) *and* (1.4). *Then*

$$\mathbb{P}\big[(\mu^\varepsilon_t)_{t\geq 0} \in \cdot\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[(\mu_t)_{t\geq 0} \in \cdot\big], \tag{1.5}$$

where $\Rightarrow$ denotes weak convergence on the Skorohod space $\mathcal{D}([0,\infty), \mathcal{M}(\mathbb{R}))$.

1.2. Main idea of the proof

We will prove Theorem 1.1 by first proving convergence of finite dimensional distributions and then proving tightness. The convergence of finite dimensional distributions will be obtained in a quick and elegant way based on results we proved earlier in [13]. Our proof of tightness is done via comparisons with the unbiased voter model, which still requires significant work and new ideas.

For each $x \in S^{01}_{\mathrm{int}}$, there exists a unique $M(x) \in \mathbb{Z} + \frac{1}{2} := \{i + \frac{1}{2} : i \in \mathbb{Z}\}$ such that

$$\sum_{i<M(x)} x(i) = \sum_{i>M(x)} \big(1 - x(i)\big), \tag{1.6}$$

that is, the number of 1's to the left of the reference point $M(x)$ equals the number of 0's to the right of it. We call $M(x)$ the *weighted midpoint* of the interface. We call

$$L(x) := \sup\left\{i \in \mathbb{Z} + \frac{1}{2} : x(j) = 0 \;\forall j < i\right\},$$
$$R(x) := \inf\left\{i \in \mathbb{Z} + \frac{1}{2} : x(j) = 1 \;\forall j > i\right\} \tag{1.7}$$

the *left* and *right boundary* of the interface, respectively. Note that $L(x) = M(x) = R(x)$ if and only if x is a *Heaviside state* of the form

$$x_{\mathrm{hv},j}(i) := 1_{\{i>j\}} \quad \left(i \in \mathbb{Z},\, j \in \mathbb{Z} + \frac{1}{2}\right). \tag{1.8}$$

In particular, we write $x_{\mathrm{hv}} := x_{\mathrm{hv},1/2}$ for the state that has 1's on the positive integers and 0's elsewhere. If x is not a Heaviside state, then $L(x) < M(x) < R(x)$. As a first step towards proving Theorem 1.1, we will prove the following weaker result.

Theorem 1.2 (Convergence of the weighted midpoint). *Fix $x \in S^{01}_{\mathrm{int}}$ and for $\varepsilon \in (0, 1)$, let X^ε be the biased voter model with generator* (1.1) *and initial state x. Then*

$$\mathbb{P}\left[\left(\varepsilon M\left(X^\varepsilon_{\varepsilon^{-2}t}\right)\right)_{t\ge 0} \in \cdot\,\right] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\left[(B_t)_{t\ge 0} \in \cdot\,\right], \tag{1.9}$$

where $\Rightarrow$ denotes weak convergence on the Skorohod space $\mathcal{D}([0,\infty), \mathbb{R})$, and $(B_t)_{t\ge 0}$ is the drifted Brownian motion defined above (1.4).

We remark that the idea of working on the weighted midpoint process is crucial. On the one hand, in the scaling limit the left and right boundaries, and hence the measure-valued process, are well approximated by the weighted midpoint in finite dimensional distributions. On the other hand, unlike the measure-valued process, there exists an efficient representation of the weighted midpoint, which is described below.

Whenever there is a site flipping from 1 (resp. 0) to 0 (resp. 1), the weighted midpoint jumps to right (resp. left) by one. Using this, it is easy to see that the weighted midpoint evolves as a random time-changed random walk, which has a drift of order ε. In view of this, to prove Theorem 1.2, it suffices to control the random time change. Somewhat surprisingly, it turns out that Lemma 3.1 and Proposition 3.7 of [13] give expressions for exactly the quantity we need and Theorem 1.2 now follows from some relatively simple renewal arguments. The paper [13] is concerned with *interface tightness*, which we explain now.

We call two configurations $x, y \in \{0, 1\}^{\mathbb{Z}}$ *equivalent*, denoted by $x \sim y$, if one is a translation of the other, that is, there exists some $k \in \mathbb{Z}$ such that $x(i) = y(i + k)$ $(i \in \mathbb{Z})$. We let $\overline{x}$ denote the equivalence class containing x and write

$$\overline{S}^{01}_{\mathrm{int}} := \left\{\overline{x} : x \in S^{01}_{\mathrm{int}}\right\}. \tag{1.10}$$

Note that S^{01}_{int} and $\overline{S}^{01}_{\mathrm{int}}$ are countable sets. Since our rates are translation invariant, the *process modulo translations* $(\overline{X}^\varepsilon_t)_{t\ge 0}$ is itself a Markov process; if we restrict the state space to $\overline{S}^{01}_{\mathrm{int}}$, then it is in fact a continuous-time Markov chain. If a is non-nearest-neighbor, then it can be shown that this Markov chain is irreducible (see [13], Lemma 2.1). In contrast, if a is nearest-neighbor, then the Markov chain is absorbed at the Heaviside state $\overline{x}_{\mathrm{hv}}$ almost surely. Following [7], we say that $(X^\varepsilon_t)_{t\ge 0}$ exhibits *interface* *tightness on* S^{01}_{int} if $(\overline{X}^{\varepsilon}_t)_{t\geq 0}$ is positive recurrent on $\overline{S}^{01}_{\mathrm{int}}$. Under our assumptions (i)–(iii) on the kernel a, interface tightness for biased voter models has been proved in [13], Thm 1.2.

Interface tightness tells us that the biased voter model, started from any initial state in S^{01}_{int}, spends a positive fraction of its time in Heaviside states. Moreover, the process modulo translations, started from $\overline{x}_{\mathrm{hv}}$, returns to $\overline{x}_{\mathrm{hv}}$ in finite expected time. Finally, the laws of the *width* of the interface $\mathbb{P}[R(X_t) - L(X_t) \in \cdot]$ are tight as $t \to \infty$. Theorem 1.3 of [13] shows that all these statements hold uniformly as the bias ε tends to zero.

Combining [13], Thms 1.2 and 1.3, which prove interface tightness uniformly as $\varepsilon \downarrow 0$, with the convergence of the weighted midpoint, we then rather easily also obtain convergence in finite dimensional distributions of the measure-valued process. To complete the proof of Theorem 1.1, it therefore suffices to show tightness of the laws of the measure-valued processes $(\mu^{\varepsilon}_t)_{t\geq 0}$ as $\varepsilon \downarrow 0$. In the unbiased setting, this has been proved in [3] by directly verifying Jakubowski's tightness criterion (see, e.g., [8], Thm 3.6.4). To use this criterion in the biased setting, we will construct a sufficient condition (2.53) in Lemma 2.13, which requires to show that over a very short time interval, there cannot be too many 0's invading far into the region dominated by 1's and vice versa. The first scenario can be ruled out by a direct comparison with the unbiased voter model. The second scenario is more subtle and can be ruled out from the observation that, once enough 1's invade far into the region dominated by 0's, they will persist till the end of the short time interval, which will contradict the already proved fact that at any deterministic time, the interface location is distributed as that of a drifted Brownian motion. Combining the two scenarios, we are able to check (2.53) and hence prove tightness.

1.3. Discussion and open problems

Combining Theorem 1.2 with [13], Thm 1.3, one can easily show that as $\varepsilon \downarrow 0$, the diffusively rescaled left boundary $(\varepsilon L(X^{\varepsilon}_{\varepsilon^{-2}t}))_{t\geq 0}$ and right boundary $(\varepsilon R(X^{\varepsilon}_{\varepsilon^{-2}t}))_{t\geq 0}$ of the interface converge in finite dimensional distributions to the same drifted Brownian motion as the weighted midpoint. A natural question then arises. That is, as $\varepsilon \downarrow 0$, do the boundaries also converge as processes, or equivalently does path level tightness for the boundaries hold?

In the unbiased case $\varepsilon = 0$, this question has been answered in a sequence of papers. Newman, Ravishankar and Sun [11] confirmed path level tightness under the assumption that a has a finite fifth moment. This result was later extended by Belhaouari et al. in [4] to all a with a finite $(3+\delta)$-th moment for some $\delta > 0$. On the other hand, it was pointed out in [4] that path level tightness for the left and right boundaries does not hold if $\sum_k a(k)|k|^{\gamma} = \infty$ for some $\gamma < 3$.

Indeed, in this regime, there exist exceptional times when 1's (resp. 0's) are created deep into the territory of the 1's (resp. 0's) du e to the heavy tail of a. Nevertheless, such 1's and 0's are expected to be rare and sparse, thanks to interface tightness. Therefore, one should be able to restore tightness if those rare 1's and 0's are suitably discounted. In [4], this idea was achieved by suppressing the infections $0 \to 1$ and $1 \to 0$ from site i to site j with $|i-j| \geq \epsilon^{-\kappa}$ for some $\kappa > 0$ depending on a, where a is required to have a finite γ-th moment for some $\gamma > 2$. The same idea also motivated Athreya and Sun [3], who proved an unbiased version of Theorem 1.1 assuming only that a has a finite second moment. We note that for interface tightness, the finite second moment assumption is optimal since it is shown in [5] that if $\sum_k a(k)|k|^{\gamma} = \infty$ for some $\gamma < 2$, then interface tightness for the (unbiased) voter model does not hold.

It is well known that the voter model is dual to a system of coalescing random walks. Likewise, the biased voter model is dual to a system of branching and coalescing random walks. It has been shown in [10] that nearest-neighbor systems of coalescing random walks, started from every point in space and time, have a diffusive scaling limit, called the *Brownian web*. Likewise, it has been shown in [12] that nearest-neighbor systems of weakly branching and coalescing random walks have a diffusive scaling limit called the *Brownian net*. To extend these results to non-nearest-neighbor systems of (branching) coalescing random walks, one needs to prove tightness for the collection of paths in the Brownian web topology, introduced in [10]. In the unbiased case, it has been shown in [4] that tightness of coalescing random walks in the Brownian web topology is equivalent to path level tightness for the left and right boundaries of the dual voter model. Their arguments carry over to branching coalescing random walks and their dual, the biased voter model.

In view of this, for the biased voter model, it is an important open problem to derive sufficient conditions for path level tightness for the left and right boundaries. We conjecture that as in the unbiased case, a finite $(3+\delta)$-th moment should suffice.

The remainder of the paper (which consists of Section 2 and an Appendix) is devoted to proofs. In Section 2.1, we give the main line of the proof of Theorem 1.2 and in Sections 2.2 and 2.3 we fill in the details. In Sections 2.4 and 2.5, we then prove Theorem 1.1 by first showing convergence in finite dimensional distributions and then tightness. Lastly, we collect some technical lemmas in the Appendix.

2. Proofs

2.1. Convergence of the weighted midpoint

In this subsection, we outline the proof of Theorem 1.2. We show that Theorem 1.2 follows from Lemmas 2.1 and 2.2 below. Here Lemma 2.1 says that the weighted midpoint evolves as a time-changed simple random walk, while Lemma 2.2 contains a statement about the convergence of the time change. We also show how Lemma 2.2 can heuristically be derived from results proved in [13], which prepares for its formal proof in Section 2.3 below.

For $x \in S^{01}_{\mathrm{int}}$ and $k \in \mathbb{Z}$, let

$$I_k(x) := \big\{i : x(i) \neq x(i+k)\big\} \tag{2.1}$$

denote the number of *k-boundaries* in the interface configuration x. Note that $I_0(x) = 0$ and $I_k(x) = I_{-k}(x)$ by definition.

Lemma 2.1 (Time-changed random walk). *Fix $x \in S^{01}_{\mathrm{int}}$ and for $\varepsilon \in (0,1)$, let X^ε be the biased voter model with generator (1.1) and initial state x. Then there exists an a.s. unique random, strictly increasing continuous function $t \mapsto T^\varepsilon_t$ such that*

$$t =: \int_0^{T^\varepsilon_t} \mathrm{d}s \sum_k a(k) I_k\big(X^\varepsilon_s\big) \quad (t \geq 0). \tag{2.2}$$

Moreover, the process $(M(X^\varepsilon_{T^\varepsilon_t}))_{t\geq 0}$ is a continuous-time Markov chain on $\mathbb{Z}+\frac{1}{2}$ that jumps as

$$m \mapsto m-1 \quad \textit{with rate } \frac{1}{2} \quad \text{and} \quad m \mapsto m+1 \quad \textit{with rate } \frac{1}{2}(1-\varepsilon). \tag{2.3}$$

By standard results, the drifted random walk in (2.3) converges after diffusive rescaling to a drifted Brownian motion. In view of this, in order to prove Theorem 1.2, the main task is to control the time-change in (2.2). Interestingly, the time-change expression also appeared in the voter model equilibrium equation (see (2.11) below), from which we can show that the time-change converges uniformly to a deterministic limit.

Lemma 2.2 (Convergence of the time change). *Let X^ε be as in Theorem 1.2. Then*

$$\sup_{0\le t\le T}\left|\sigma^2 t-\varepsilon^2\int_0^{\varepsilon^{-2}t}\mathrm{d}s\sum_k a(k)I_k\big(X_s^\varepsilon\big)\right|\xrightarrow[\varepsilon\to 0]{\mathrm{P}}0\quad(T<\infty),\tag{2.4}$$

where $\xrightarrow{\mathrm{P}}$ denotes convergence in probability.

Proof of Theorem 1.2. Let us set

$$Y_t^\varepsilon:=\varepsilon M\big(X^\varepsilon_{T^\varepsilon_{\varepsilon^{-2}t}}\big)\quad(t\ge 0),\tag{2.5}$$

i.e., this is the drifted random walk $(M(X^\varepsilon_{T_t^\varepsilon}))_{t\ge 0}$ from Lemma 2.1, diffusively rescaled by ε. Then standard results tell us that

$$\mathbb{P}\big[\big(Y_t^\varepsilon\big)_{t\ge 0}\in\cdot\,\big]\underset{\varepsilon\to 0}{\Longrightarrow}\mathbb{P}\Big[\Big(W_t-\frac{1}{2}t\Big)_{t\ge 0}\in\cdot\,\Big],\tag{2.6}$$

where $(W_t)_{t\ge 0}$ is a standard Brownian motion. Let

$$S_t^\varepsilon=\int_0^t\mathrm{d}s\sum_k a(k)I_k\big(X_s^\varepsilon\big)\quad(t\ge 0)\tag{2.7}$$

denote the inverse of the function $t\mapsto T_t^\varepsilon$ defined in (2.2). Then

$$\varepsilon M\big(X^\varepsilon_{\varepsilon^{-2}t}\big)=Y^\varepsilon_{\varepsilon^2 S^\varepsilon_{\varepsilon^{-2}t}}\quad(t\ge 0).\tag{2.8}$$

Lemma 2.2 tells us that $\varepsilon^2 S^\varepsilon_{\varepsilon^{-2}t}$ converges as a process to $\sigma^2 t$. It is not hard to show (for details, we refer to Lemma A.5 in the Appendix) that this implies convergence of the time-changed process, proving the claim of Theorem 1.2. □

In order to prove the crucial Lemma 2.2, we will rely on results proved in [13]. In the remainder of this subsection, we recall some of these results and put them into context, to give the reader a rough idea where Lemma 2.2 comes from.

As explained in Section 1.2, Theorem 1.2 in [13] establishes interface tightness for biased voter models. More precisely, this theorem says that for any $\varepsilon\in[0,1)$, the process modulo translations $(\overline{X}_t^\varepsilon)_{t\ge 0}$ has a unique invariant law. Let us denote this invariant law by $\overline{\pi}^\varepsilon$. In particular, if a is nearest-neighbor (and therefore $a(-1)=\frac{1}{2}=a(1)$ by our assumptions on a), then the Heaviside state $\overline{x}_{\mathrm{hv}}$ is absorbing and $\overline{\pi}^\varepsilon$ is the delta measure on $\overline{x}_{\mathrm{hv}}$. In the non-nearest-neighbor case, we cite the following theorem from [13], Thm 1.3. The extension to the nearest-neighbor case is trivial.

Theorem 2.3 (Continuity of the invariant law). *The laws $\overline{\pi}^\varepsilon$ converge weakly to $\overline{\pi}^0$ as $\varepsilon\downarrow 0$ with respect to the discrete topology on $\overline{S}^{01}_{\mathrm{int}}$.*

All existing proofs of interface tightness for unbiased voter models are in some way or another based on a function that counts the number of *inversions*, that is, pairs of sites i,j such that $i<j$ and $x(i)>x(j)$. Let h denote this function, that is,

$$h(x):=\sum_{i<j}1_{\{x(i,j)=10\}}\quad\big(x\in S^{01}_{\mathrm{int}}\big).\tag{2.9}$$

Note that since h is translation invariant, we can alternatively view h as a function on $\overline{S}^{01}_{\text{int}}$. In [13], Prop. 3.7, it is shown that the invariant law $\overline{\pi}^0$ solves the equilibrium equation

$$\sum_{\overline{x}\in\overline{S}^{01}_{\text{int}}} \overline{\pi}^0(\overline{x}) G^0 h(\overline{x}) = 0, \tag{2.10}$$

where G^0 is the generator defined in (1.1). As shown in [13], Prop. 3.7, this equation can be written more explicitly as follows. (In the nearest-neighbor case, [13], Prop. 3.7, is not applicable, but (2.11) below holds trivially with both sides equal to 1.)

Proposition 2.4 (Equilibrium equation). *Let X^0_∞ be a random variable such that $\overline{X}^0_\infty$ has law $\overline{\pi}^0$. Then*

$$\mathbb{E}\Big[\sum_k a(k) I_k\big(X^0_\infty\big)\Big] = \sigma^2. \tag{2.11}$$

It is a remarkable fact that the equilibrium equation for the function in (2.9) yields an expression for precisely the quantity that also appears in the time-change in (2.2). Proposition 2.4 was one of the main ingredients used in [13] to prove Theorem 2.3. Together, Theorem 2.3 and Proposition 2.4 will be the main ingredients in our proof of Lemma 2.2. In order to derive Lemma 2.2 from (2.11), we will need uniform control over the speed at which the process modulo translations converges to equilibrium. This will be achieved by a renewal decomposition of the process modulo translations, where we get uniform control on the expected return times as $\varepsilon \downarrow 0$ as a result of Theorem 2.3.

In the coming two subsections, we prove Lemmas 2.1 and 2.2, respectively.

2.2. Random time-change

Proof of Lemma 2.1. Let

$$\begin{aligned} I^{01}_k(x) &:= \big\{i : x(i) = 0, x(i+k) = 1\big\}, \\ I^{10}_k(x) &:= \big\{i : x(i) = 1, x(i+k) = 0\big\}. \end{aligned} \tag{2.12}$$

Since for every $x \in S^{01}_{\text{int}}$, $k > 0$ and $i \in \mathbb{Z}$, there is one more adjacent pair of (01) than (10) along the subsequence $x(\dots, i-k, i, i+k, \dots)$, it is not hard to see that

$$I^{01}_k(x) = I^{10}_k(x) + k \quad \text{and} \quad I_k(x) = I^{01}_k(x) + I^{10}_k(x) \quad \big(x \in S^{01}_{\text{int}}\big). \tag{2.13}$$

As a result

$$I^{10}_k(x) = \frac{1}{2}\big(I_k(x) - k\big) \quad \text{and} \quad I^{01}_k(x) = \frac{1}{2}\big(I_k(x) + k\big). \tag{2.14}$$

We observe that the quantity $M(X^\varepsilon_t)$ always goes up and down by a single unit. More precisely, $M(X^\varepsilon_t)$ goes down by one when a site flips from 0 to 1 and it goes up by one when a site flips from 1 to 0, which means that if the present state is $X^\varepsilon_t = x$, then $M(X^\varepsilon_t)$ jumps as

$$\begin{aligned} m &\mapsto m-1 \quad \text{with rate } \sum_k a(k) I^{10}_k(x) = \frac{1}{2}\sum_k a(k) I_k(x), \\ m &\mapsto m+1 \quad \text{with rate } (1-\varepsilon)\sum_k a(k) I^{01}_k(x) = (1-\varepsilon)\frac{1}{2}\sum_k a(k) I_k(x), \end{aligned} \tag{2.15}$$

where we have used (2.14) and our assumption $\sum_k a(k)k = 0$. It follows from (2.15) that $M(X_t^\varepsilon)$ is a drifted random walk with a random time change.

More precisely, defining S_t^ε as in (2.7), and observing that the integrand is ≥ 1, we see that S_t^ε is a.s. strictly increasing, continuous, with $S_0^\varepsilon = 0$ and $\lim_{t\to\infty} S_t^\varepsilon = \infty$. It follows that S_t^ε has an inverse function with the same properties, which is T_t^ε. By standard results, the time-changed process $(X^\varepsilon_{T_t^\varepsilon})_{t\geq 0}$ is a Markov process such that if the original process $(X_t^\varepsilon)_{t\geq 0}$ jumps from x to y with rate $r(x,y)$, then the new process $(X^\varepsilon_{T_t^\varepsilon})_{t\geq 0}$ jumps from x to y with rate $(\sum_k a(k)I_k(x))^{-1} r(x,y)$. In particular, the process $(M(X^\varepsilon_{T_t^\varepsilon}))_{t\geq 0}$ is a drifted random walk with jump rates as in (2.3). □

2.3. Renewal arguments

In this subsection, we prove Lemma 2.2, completing the proof of Theorem 1.2. Since the functions I_k are translation invariant, we can and will view them as functions on $\overline{S}^{01}_{\mathrm{int}}$. Our task is then to show that if $\overline{x} \in \overline{S}^{01}_{\mathrm{int}}$ is fixed and $\overline{X}^\varepsilon$ is the biased voter model with bias $\varepsilon \in (0,1)$, modulo translations, started in $\overline{x}$, then

$$\sup_{0\leq t\leq T} \left| \sigma^2 t - \varepsilon^2 \int_0^{\varepsilon^{-2}t} \mathrm{d}s \sum_k a(k) I_k\big(\overline{X}_s^\varepsilon\big) \right| \xrightarrow[\varepsilon\to 0]{\mathrm{P}} 0. \tag{2.16}$$

We let

$$\tau_0^\varepsilon := \inf\big\{t \geq 0 : \overline{X}_t^\varepsilon = \overline{x}_{\mathrm{hv}}\big\} \tag{2.17}$$

denote the first hitting time of $\overline{x}_{\mathrm{hv}}$, and define inductively

$$\tilde{\tau}_n^\varepsilon := \inf\big\{t > \tau_{n-1}^\varepsilon : \overline{X}_t^\varepsilon \neq \overline{x}_{\mathrm{hv}}\big\} \quad \text{and} \quad \tau_n^\varepsilon := \inf\big\{t > \tilde{\tau}_n^\varepsilon : \overline{X}_t^\varepsilon = \overline{x}_{\mathrm{hv}}\big\} \quad (n \geq 1). \tag{2.18}$$

We will first prove (2.16) under the additional assumptions that $\overline{X}_0^\varepsilon = \overline{x}_{\mathrm{hv}}$ and the kernel a is non-nearest-neighbor. The assumption that $\overline{X}_0^\varepsilon = \overline{x}_{\mathrm{hv}}$ implies that τ_0^ε from (2.17) is zero, while the assumption that a is non-nearest-neighbor implies that $r_\varepsilon > 0$, where

$$r_\varepsilon := \sum_{k<-1} \big(|k|-1\big)a(k) + (1-\varepsilon)\sum_{k>1}(k-1)a(k) \tag{2.19}$$

is the rate at which $\overline{X}^\varepsilon$ jumps away from $\overline{x}_{\mathrm{hv}}$. We start with a trivial observation. Below, we view the law of $\overline{X}^\varepsilon$ as a probability measure on the space of piecewise constant, right-continuous functions with values in the countable set $\overline{S}^{01}_{\mathrm{int}}$, and we equip this space with the Skorohod topology.

Lemma 2.5 (Continuity of the law). *Let $\overline{X}^\varepsilon = (\overline{X}_t^\varepsilon)_{t\geq 0}$ be the biased voter model modulo translations with bias $\varepsilon \in [0,1)$, started in $\overline{X}_0^\varepsilon = \overline{x}_{\mathrm{hv}}$. Then the function $\varepsilon \mapsto \mathbb{P}[\overline{X}^\varepsilon \in \cdot\,]$ is continuous with respect to weak convergence.*

Proof. This is trivial, since $\overline{X}^\varepsilon$ is a nonexplosive continuous-time Markov chain and its jump rates converge pointwise. □

Lemma 2.6 (Convergence of return times). *Assume that a is non-nearest-neighbor. Then*

$$\lim_{\varepsilon\downarrow 0} \mathbb{E}^{\overline{x}_{\mathrm{hv}}}\big[\tau_1^\varepsilon\big] = \mathbb{E}^{\overline{x}_{\mathrm{hv}}}\big[\tau_1^0\big] < \infty. \tag{2.20}$$

Proof. By interface tightness [13], Thm 1.2, we have $\mathbb{E}[\tau_1^\varepsilon] < \infty$ for each $\varepsilon \in [0,1)$. The regenerative theorem (see [2], Thm 4.1.2) gives an expression for the invariant law $\overline{\pi}^\varepsilon$,

$$\overline{\pi}^\varepsilon(\overline{x}) = \frac{1}{\mathbb{E}^{\overline{x}_{\mathrm{hv}}}[\tau_1^\varepsilon]}\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\left[\int_0^{\tau_1^\varepsilon} 1_{\{\overline{X}_s^\varepsilon = \overline{x}\}}\,\mathrm{d}s\right] \quad \big(\overline{x} \in \overline{S}_{\mathrm{int}}^{01}\big). \tag{2.21}$$

In particular, setting $\overline{x} = \overline{x}_{\mathrm{hv}}$ and using the fact that during $[0,\tau_1^\varepsilon)$ one has $\overline{X}_s^\varepsilon = \overline{x}_{\mathrm{hv}}$ if and only if $s \in [0,\tilde{\tau}_1^\varepsilon)$, it follows that

$$\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\big[\tau_1^\varepsilon\big] = \frac{1}{\overline{\pi}^\varepsilon(\overline{x}_{\mathrm{hv}})}\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\big[\tilde{\tau}_1^\varepsilon\big] = \frac{1}{r_\varepsilon \overline{\pi}^\varepsilon(\overline{x}_{\mathrm{hv}})}, \tag{2.22}$$

where r_ε from (2.19) is the rate at which $\overline{X}^\varepsilon$ leaves $\overline{x}_{\mathrm{hv}}$. By Theorem 2.3, $\overline{\pi}^\varepsilon(\overline{x}_{\mathrm{hv}}) \to \overline{\pi}^0(\overline{x}_{\mathrm{hv}})$ as $\varepsilon \to 0$, which together with (2.22) yields the claim. □

Lemma 2.7 (Average value during one excursion). *Assume that a is non-nearest-neighbor. Then*

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\left[\int_0^{\tau_1^\varepsilon}\mathrm{d}s \sum_k a(k) I_k\big(\overline{X}_s^\varepsilon\big)\right] = \mathbb{E}^{\overline{x}_{\mathrm{hv}}}\left[\int_0^{\tau_1^0}\mathrm{d}s \sum_k a(k) I_k\big(\overline{X}_s^0\big)\right] < \infty. \tag{2.23}$$

Proof. Formula (2.21) gives

$$\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\left[\int_0^{\tau_1^\varepsilon}\mathrm{d}s \sum_k a(k) I_k\big(\overline{X}_s^\varepsilon\big)\right] = \mathbb{E}^{\overline{x}_{\mathrm{hv}}}\big[\tau_1^\varepsilon\big] \sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^\varepsilon(\overline{x}) \sum_k a(k) I_k(\overline{x}). \tag{2.24}$$

By Theorem 2.3 and Fatou's lemma,

$$\sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^0(\overline{x}) \sum_k a(k) I_k(\overline{x}) \le \liminf_{\varepsilon\downarrow 0} \sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^\varepsilon(\overline{x}) \sum_k a(k) I_k(\overline{x}). \tag{2.25}$$

By [13], Lemma 3.1, and Proposition 2.4,

$$\sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^\varepsilon(\overline{x}) \sum_k a(k) I_k(\overline{x}) \le \sigma^2 \quad \text{and} \quad \sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^0(\overline{x}) \sum_k a(k) I_k(\overline{x}) = \sigma^2. \tag{2.26}$$

The right-hand side of (2.25) is less than or equal to the limit superior of the same expression, which by the first formula in (2.26) is bounded from above by σ^2, while the second formula in (2.26) identifies the left-hand side of (2.25) as σ^2. We conclude that

$$\sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^\varepsilon(\overline{x}) \sum_k a(k) I_k(\overline{x}) \underset{\varepsilon\to 0}{\longrightarrow} \sigma^2 = \sum_{\overline{x}\in\overline{S}_{\mathrm{int}}^{01}} \overline{\pi}^0(\overline{x}) \sum_k a(k) I_k(\overline{x}). \tag{2.27}$$

Inserting this into (2.24), using Lemma 2.6, we obtain (2.23). □

The proof of Lemma 2.7 yields a useful corollary.

Corollary 2.8 (Renewal identity). *Assume that a is non-nearest-neighbor. Then*

$$\frac{1}{\mathbb{E}^{\overline{x}_{\mathrm{hv}}}[\tau_1^0]}\mathbb{E}^{\overline{x}_{\mathrm{hv}}}\Big[\int_0^{\tau_1^0}\mathrm{d}s\sum_k a(k)I_k\big(\overline{X}_s^0\big)\Big]=\sigma^2. \tag{2.28}$$

Proof. This follows from (2.24) and (2.26). □

Let $\overline{X}^\varepsilon$ denote the process started in $\overline{X}_0^\varepsilon=\overline{x}_{\mathrm{hv}}$ and let

$$\phi_\varepsilon(u):=\varepsilon^2\sum_{k=1}^{\lfloor\varepsilon^{-2}u\rfloor}\big(\tau_k^\varepsilon-\tau_{k-1}^\varepsilon\big)\quad\text{and}\quad\psi_\varepsilon(u):=\varepsilon^2\sum_{k=1}^{\lfloor\varepsilon^{-2}u\rfloor}\int_{\tau_{k-1}^\varepsilon}^{\tau_k^\varepsilon}\mathrm{d}s\sum_k a(k)I_k\big(\overline{X}_s^\varepsilon\big). \tag{2.29}$$

By the strong Markov property, $\phi_\varepsilon(u)$ and $\psi_\varepsilon(u)$ are sums of i.i.d. random variables. Indeed, $\tau_k^\varepsilon-\tau_{k-1}^\varepsilon$ is equally distributed with τ_1^ε while the summands of $\psi_\varepsilon(u)$ are equally distributed with

$$\eta^\varepsilon:=\int_0^{\tau_1^\varepsilon}\mathrm{d}s\sum_k a(k)I_k\big(\overline{X}_s^\varepsilon\big). \tag{2.30}$$

It follows from Lemma 2.5 that τ_1^ε and η^ε converge weakly in law as $\varepsilon\to 0$ to τ_1^0 and η^0, respectively. Note that $\tau_1^0,\eta^0>0$ a.s. Lemmas 2.6, 2.7, and Corollary 2.8 tell us that

$$\lim_{\varepsilon\to 0}\mathbb{E}\big[\tau_1^\varepsilon\big]=\mathbb{E}\big[\tau_1^0\big]<\infty,\qquad\lim_{\varepsilon\to 0}\mathbb{E}\big[\eta^\varepsilon\big]=\mathbb{E}\big[\eta^0\big]<\infty,\quad\text{and}\quad\mathbb{E}\big[\eta^0\big]/\mathbb{E}\big[\tau_1^0\big]=\sigma^2. \tag{2.31}$$

By [9], Prop. A.2.3, the convergence in law and in expectation of τ_1^ε and η^ε imply that these random variables are uniformly integrable as $\varepsilon\downarrow 0$, i.e., for any $\varepsilon_n\to 0$, we have

$$\lim_{K\to\infty}\sup_n\mathbb{E}\big[\tau_1^{\varepsilon_n};\tau_1^{\varepsilon_n}\ge K\big]=0\quad\text{and}\quad\lim_{K\to\infty}\sup_n\mathbb{E}\big[\eta^{\varepsilon_n};\eta^{\varepsilon_n}\ge K\big]=0. \tag{2.32}$$

It follows that

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[\tau_1^\varepsilon;\tau_1^\varepsilon>t\varepsilon^{-2}\big]=0\quad\text{and}\quad\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[\eta^\varepsilon;\eta^\varepsilon>t\varepsilon^{-2}\big]=0\quad(t>0). \tag{2.33}$$

This allows us to apply a standard functional law of large numbers (see Lemma A.8 in the Appendix for details) to obtain the following lemma.

Lemma 2.9 (Functional law of large numbers). *Let ϕ_ε and ψ_ε be as in* (2.29). *Then*

$$\sup_{0\le u\le U}\big|u\mathbb{E}\big[\tau_1^0\big]-\phi_\varepsilon(u)\big|\underset{\varepsilon\to 0}{\overset{\mathrm{P}}{\longrightarrow}}0\quad\text{and}\quad\sup_{0\le u\le U}\big|u\mathbb{E}\big[\eta^0\big]-\psi_\varepsilon(u)\big|\underset{\varepsilon\to 0}{\overset{\mathrm{P}}{\longrightarrow}}0\quad(U>0). \tag{2.34}$$

Proof of Lemma 2.2. We will finish the proof in two steps, first assuming a Heaviside initial state and then extending to arbitrary initial states.

Heaviside initial state. We first prove the statement under the additional assumptions that the kernel a is non-nearest-neighbor and $\overline{X}_0^\varepsilon=\overline{x}_{\mathrm{hv}}$. In this case, $\tau_0^\varepsilon=0$.

Since $\phi_\varepsilon(\varepsilon^2k)=\varepsilon^2\tau_k^\varepsilon$ $(k\ge 0)$, the function ϕ_ε defines a bijection from $\varepsilon^2\mathbb{N}$ to $\{\varepsilon^2\tau_k^\varepsilon:k\ge 0\}$. Let θ_ε denote the restriction of ϕ_ε to $\varepsilon^2\mathbb{N}$. Then ϕ_ε is the right-continuous interpolation of θ_ε. Let θ_ε^{-1} denote the inverse of θ_ε. For any $t \geq 0$ and $k \in \mathbb{N}$, define

$$[t]^\varepsilon_\leftarrow := \tau^\varepsilon_{k-1} \quad \text{where } t \in \big[\tau^\varepsilon_{k-1}, \tau^\varepsilon_k\big) \quad \text{and} \quad [t]^\varepsilon_\rightarrow := \tau^\varepsilon_k \quad \text{where } t \in \big(\tau^\varepsilon_{k-1}, \tau^\varepsilon_k\big]. \tag{2.35}$$

Then for $t \geq 0$,

$$\psi_\varepsilon\big(\theta_\varepsilon^{-1}\big(\varepsilon^2\big[\varepsilon^{-2}t\big]^\varepsilon_\leftarrow\big)\big) \leq \varepsilon^2 \int_0^{\varepsilon^{-2}t} \mathrm{d}s \sum_k a(k) I_k\big(\overline{X}^\varepsilon_s\big) \leq \psi_\varepsilon\big(\theta_\varepsilon^{-1}\big(\varepsilon^2\big[\varepsilon^{-2}t\big]^\varepsilon_\rightarrow\big)\big), \tag{2.36}$$

where $\theta_\varepsilon^{-1}(\varepsilon^2[\varepsilon^{-2}t]^\varepsilon_\leftarrow)$ and $\theta_\varepsilon^{-1}(\varepsilon^2[\varepsilon^{-2}t]^\varepsilon_\rightarrow)$ are the right- and left-continuous interpolations of the function θ_ε^{-1}.

Lemma 2.9 tells us that as $\varepsilon \downarrow 0$, the right-continuous interpolation of θ_ε converges in probability w.r.t. the Skorohod topology to the function $u \mapsto u\mathbb{E}[\tau_1^0]$. By the Skorohod representation theorem [6], Thm 6.7, along any sequence $\varepsilon_n \downarrow 0$, we can couple our random variables such that this convergence is a.s. Since θ_ε takes values in $\{\varepsilon^2\tau^\varepsilon_k : k \geq 0\}$, it is easy to see that for our coupling

$$\forall t \geq 0 \quad \exists t_n \in \big\{\varepsilon_n^2 \tau_k^{\varepsilon_n} : k \geq 0\big\} \quad \text{s.t.} \quad t_n \to t, \tag{2.37}$$

that is, the range of θ_{ε_n} is a.s. dense in the limit. Since the sets $\varepsilon_n^2\mathbb{N}$ are dense in the limit, it is easy to see that not only the right-continuous interpolation, but also the linear interpolation of θ_{ε_n} converges locally uniformly to the function $u \mapsto u\mathbb{E}[\tau_1^0]$. It is not hard to see that this implies locally uniform convergence of the inverse (see Lemma A.4 in the Appendix). Thus, the linear interpolation of $\theta_{\varepsilon_n}^{-1}$ converges locally uniformly to the function $t \mapsto t/\mathbb{E}[\tau_1^0]$. Using (2.37), we see that the same holds for the right- and left-continuous interpolations of the function $\theta_{\varepsilon_n}^{-1}$. Since this holds for arbitrary $\varepsilon_n \downarrow 0$, we obtain that

$$\sup_{0 \leq t \leq T} \big|t/\mathbb{E}\big[\tau_1^0\big] - \theta_\varepsilon^{-1}\big(\varepsilon^2\big[\varepsilon^{-2}t\big]^\varepsilon_\leftarrow\big)\big| \xrightarrow[\varepsilon \to 0]{\mathrm{P}} 0, \tag{2.38}$$

and similarly for the left-continuous interpolation. Combining this with Lemma 2.9, it is not hard to show (see Lemma A.5 from the Appendix) that the left- and right-hand sides of (2.36) converge locally uniformly in probability to the composition of the functions $t \mapsto t/\mathbb{E}[\tau_1^0]$ and $u \mapsto u\mathbb{E}[\eta^0]$. By (2.31), this composite function is the function $t \mapsto \sigma^2 t$, proving (2.4).

This completes the proof under the additional assumptions that the kernel a is non-nearest-neighbor and $\overline{X}^\varepsilon_0 = \overline{x}_{\mathrm{hv}}$. If a is the nearest-neighbor kernel $a(-1) = \frac{1}{2} = a(1)$, then $\sigma^2 = 1$ and $\overline{X}^\varepsilon_t = \overline{x}_{\mathrm{hv}}$ for each $t \geq 0$. Moreover, $\sum_k a(k) I_k(\overline{x}_{\mathrm{hv}}) = 1$, so in this case (2.16) is trivial.

Arbitrary initial states. To treat the case when $\overline{X}^\varepsilon$ started in an arbitrary, fixed initial state $\overline{X}^\varepsilon_0 = \overline{x}$, it suffices to show that

$$\varepsilon^2 \tau_0^\varepsilon \xrightarrow[\varepsilon \to 0]{\mathrm{P}} 0 \quad \text{and} \quad \varepsilon^2 \int_0^{\tau_0^\varepsilon} \mathrm{d}s \sum_k a(k) I_k\big(\overline{X}^\varepsilon_s\big) \xrightarrow[\varepsilon \to 0]{\mathrm{P}} 0. \tag{2.39}$$

Since the jump rates converge, for any $\overline{x} \in \overline{S}^{01}_{\mathrm{int}}$, the laws

$$\mathbb{P}^{\overline{x}}\big[\tau_0^\varepsilon \in \cdot\big] \quad \text{and} \quad \mathbb{P}^{\overline{x}}\bigg[\int_0^{\tau_0^\varepsilon} \mathrm{d}s \sum_k a(k) I_k\big(\overline{X}^\varepsilon_s\big) \in \cdot\bigg] \tag{2.40}$$

converge weakly as $\varepsilon \downarrow 0$, so it suffices to show that for the unbiased process

$$\mathbb{P}^{\overline{x}}\big[\tau_0^0 < \infty\big] = 1 \quad \text{and} \quad \mathbb{P}^{\overline{x}}\bigg[\int_0^{\tau_0^0} \mathrm{d}s \sum_k a(k) I_k\big(\overline{X}_s^0\big) < \infty\bigg] = 1, \tag{2.41}$$

where in fact the first equality implies the second equality. If the kernel a is non-nearest-neighbor, then the fact that $\tau_0^0 < \infty$ for any $\overline{x} \in \overline{S}_{\text{int}}^{01}$ follows from positive recurrence and irreducibility [13], Thm 1.2 and Lemma 2.1.

To complete the proof, we must show that the nearest-neighbor unbiased voter model, started in any state $x \in S_{\text{int}}^{01}$, a.s. reaches a Heaviside state in finite time. We can obtain a two-type voter model as a function of a multi-type voter model in which initially each site has a different type. Since the family size of each type is a martingale, each family dies out a.s. As soon as the families corresponding to types that were initially between $L(x)$ and $R(x)$ have all died out, X_t and thus $\overline{X}_t$ will be in a Heaviside state. □

2.4. Convergence of finite dimensional distributions

In this subsection, we start proving Theorem 1.1 by showing convergence in finite dimensional distributions.

Lemma 2.10 (Local limit). *Fix $\overline{x} \in \overline{S}_{\text{int}}^{01}$ and for $\varepsilon \in [0, 1)$, let $\overline{X}^{\varepsilon}$ be the biased voter model modulo translations with bias ε, started in $\overline{x}$. Then, for each $\varepsilon_n \to 0$ and $t_n \to \infty$,*

$$\mathbb{P}^{\overline{x}}\big[\overline{X}_{t_n}^{\varepsilon_n} \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \overline{\pi}^0, \tag{2.42}$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\overline{S}_{\text{int}}^{01}$ with respect to the discrete topology, and $\overline{\pi}^0$ is the invariant law of $\overline{X}^0$.

Proof. We first prove the statement if the kernel a is non-nearest-neighbor. The process $\overline{X}^{\varepsilon}$ is irreducible (by [13], Lemma 2.1) and positive recurrent (by [13], Thm 1.2) for each $\varepsilon \geq 0$. Moreover, its jump rates and by Theorem 2.3 also its invariant law converge to those of $\overline{X}^0$ as $\varepsilon \downarrow 0$. Using this, a simple abstract argument (see Lemma A.9 in the Appendix) gives

$$\sup_{n\geq 0} \big\|\mathbb{P}^{\overline{x}}\big[\overline{X}_t^{\varepsilon_n} \in \cdot\,\big] - \overline{\pi}^{\varepsilon_n}\big\| \underset{t\to\infty}{\longrightarrow} 0, \tag{2.43}$$

where $\|\cdot\|$ denotes the total variation norm. Since

$$\big\|\mathbb{P}^{\overline{x}}\big[\overline{X}_{t_n}^{\varepsilon_n} \in \cdot\,\big] - \overline{\pi}^0\big\| \leq \big\|\mathbb{P}^{\overline{x}}\big[\overline{X}_{t_n}^{\varepsilon_n} \in \cdot\,\big] - \overline{\pi}^{\varepsilon_n}\big\| + \big\|\overline{\pi}^{\varepsilon_n} - \overline{\pi}^0\big\|, \tag{2.44}$$

the claim follows from the convergence of $\overline{\pi}^{\varepsilon_n}$ (Theorem 2.3).

Since for the nearest-neighbor kernel $a(-1) = \frac{1}{2} = a(1)$, the invariant law $\overline{\pi}^0$ is the delta measure on $\overline{x}_{\text{hv}}$, in this case it suffices to prove that

$$\mathbb{P}^{\overline{x}}\big[\tau_0^{\varepsilon_n} > t_n\big] \underset{n\to\infty}{\longrightarrow} 0, \tag{2.45}$$

where $\tau_0^{\varepsilon_n}$ as in (2.17) denotes the time $\overline{X}^{\varepsilon_n}$ gets trapped in $\overline{x}_{\mathrm{hv}}$. It has already been shown below (2.40) that

$$\lim_{\varepsilon\downarrow 0}\mathbb{P}^{\overline{x}}\big[\tau_0^{\varepsilon}>t\big]=\mathbb{P}^{\overline{x}}\big[\tau_0^{0}>t\big]\quad\text{and}\quad\lim_{t\to\infty}\mathbb{P}^{\overline{x}}\big[\tau_0^{0}>t\big]=0,\tag{2.46}$$

which implies (2.45). □

Proposition 2.11 (Convergence of the left and right boundaries). *Fix $x\in S^{01}_{\mathrm{int}}$ and for $\varepsilon\in(0,1)$, let X^ε be the biased voter model with generator (1.1) and initial state x. Then*

$$\mathbb{P}\big[\big(\varepsilon L\big(X^\varepsilon_{\varepsilon^{-2}t}\big),\varepsilon R\big(X^\varepsilon_{\varepsilon^{-2}t}\big)\big)_{t\geq 0}\in\cdot\big]\underset{\varepsilon\to 0}{\overset{\text{f.d.d.}}{\Longrightarrow}}\mathbb{P}\big[(B_t,B_t)_{t\geq 0}\in\cdot\big],\tag{2.47}$$

where $\overset{\text{f.d.d.}}{\Longrightarrow}$ denotes weak convergence of the finite dimensional distributions, and $(B_t)_{t\geq 0}$ is a Brownian motion with drift $-\frac{1}{2}\sigma^2$ and diffusion coefficient σ^2.

Proof. Let $W(x):=R(x)-L(x)$ denote the width of an interface $x\in S^{01}_{\mathrm{int}}$, which by translation invariance we can view as a function on $\overline{S}^{01}_{\mathrm{int}}$. Lemma 2.10 shows that as $\varepsilon\downarrow 0$, the law of $W(\overline{X}^\varepsilon_{\varepsilon^{-2}t})$ converges to a limit law on $\mathbb{N}$, and hence

$$\mathbb{P}^x\big[\varepsilon W\big(X^\varepsilon_{\varepsilon^{-2}t}\big)\in\cdot\big]\underset{\varepsilon\to 0}{\Longrightarrow}0\quad(t>0).\tag{2.48}$$

Since x is fixed, this trivially also holds for $t=0$. By Theorem 1.2 and the Skorohod representation theorem [6], Thm 6.7, along any sequence $\varepsilon_n\downarrow 0$, we can couple our processes such that $(\varepsilon_n M(X^{\varepsilon_n}_{\varepsilon_n n^{-2}t}))_{t\geq 0}$ converges a.s. to $(B_t)_{t\geq 0}$. The claim now follows since $|L(x)-M(x)|\leq W(x)$ and similarly for $R(x)$ $(x\in S^{01}_{\mathrm{int}})$. □

Lemma 2.12 (Convergence of finite dimensional distributions). *Fix $x\in S^{01}_{\mathrm{int}}$ and for $\varepsilon\in(0,1)$, let X^ε be the biased voter model with generator (1.1) and initial state x. Define $(\mu^\varepsilon_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ as in (1.3) and (1.4). Then for each $0\leq t_1<\cdots<t_m$,*

$$\mathbb{P}\big[\big(\mu^\varepsilon_{t_1},\ldots,\mu^\varepsilon_{t_m}\big)\in\cdot\big]\underset{\varepsilon\to 0}{\Longrightarrow}\mathbb{P}\big[(\mu_{t_1},\ldots,\mu_{t_m})\in\cdot\big],\tag{2.49}$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{M}(\mathbb{R})^m$, and $\mathcal{M}(\mathbb{R})$ is the space of locally finite measures on $\mathbb{R}$, equipped with the topology of vague convergence.

Proof. Let

$$\mu^{\mathrm{l},\varepsilon}_t:=\sum_{i>L(X^\varepsilon_{\varepsilon^{-2}t})}\varepsilon\delta_{\varepsilon i}\quad\text{and}\quad\mu^{\mathrm{r},\varepsilon}_t:=\sum_{i>R(X^\varepsilon_{\varepsilon^{-2}t})}\varepsilon\delta_{\varepsilon i}.\tag{2.50}$$

By Proposition 2.11 and the Skorohod representation theorem [6], Thm 6.7, along any sequence $\varepsilon_n\downarrow 0$, we can couple our processes such that

$$\varepsilon_n L\big(X^{\varepsilon_n}_{\varepsilon_n^{-2}t_k}\big)\underset{n\to\infty}{\longrightarrow}B_{t_k}\quad\text{and}\quad\varepsilon_n R\big(X^{\varepsilon_n}_{\varepsilon_n^{-2}t_k}\big)\underset{n\to\infty}{\longrightarrow}B_{t_k}\quad\text{a.s. }(1\leq k\leq m).\tag{2.51}$$

Then, for any continuous function $f:\mathbb{R}\to\mathbb{R}$ with compact support

$$\int_{\mathbb{R}}\mu^{\mathrm{l},\varepsilon_n}_{t_k}(\mathrm{d}r)f(r)\underset{n\to\infty}{\longrightarrow}\int_{\mathbb{R}}\mu_{t_k}(\mathrm{d}r)f(r)\quad\text{a.s. }(1\leq k\leq m),\tag{2.52}$$

and similarly for $\mu_{t_k}^{\mathrm{r},\varepsilon_n}$, so using the fact that $\mu_t^{\mathrm{r},\varepsilon} \le \mu_t^{\varepsilon} \le \mu_t^{\mathrm{l},\varepsilon}$, we see that (2.52) holds with $\mu_{t_k}^{\mathrm{l},\varepsilon_n}$ replaced by $\mu_{t_k}^{\varepsilon_n}$, first for $f \ge 0$ and then for general f by linearity. This proves that $\mu_{t_k}^{\varepsilon_n}$ converges a.s. vaguely to μ_{t_k} for each $1 \le k \le m$. Since this holds for arbitrary $\varepsilon_n \downarrow 0$, (2.49) follows. $\square$

2.5. Tightness

In this subsection, we complete the proof of Theorem 1.1 by showing tightness. Roughly speaking, we need to show that over a very short time interval, there cannot be too many 0's invading far into the region dominated by 1's and vice versa. The first scenario will be ruled out by a direct comparison with the unbiased voter model, for which tightness was proved in [3]. To rule out the second scenario, the key observation is that once enough 1's invade far into the region dominated by 0's, they will persist, leading to a large displacement of the interface at the end of the short time interval.

We let $\langle \mu, \phi \rangle := \int \phi \, \mathrm{d}\mu$ denote the integral of a function ϕ with respect to a measure μ, and we write $\mathcal{C}_{\mathrm{c}}^2(\mathbb{R})$ for the space of compactly supported, twice continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Let $\mathcal{K}$ denote the space of all measures $\mu \in \mathcal{M}(\mathbb{R})$ such that $\mu([-n, n]) \le 2n + 1$ for $n = 1, 2, \ldots$. Then $\mathcal{K}$ is a compact subset of $\mathcal{M}(\mathbb{R})$ and $\mu_t^{\varepsilon} \in \mathcal{K}$ for all $\varepsilon \in [0, 1)$ and $t \ge 0$. In view of this, by Jakubowski's tightness criterion [8], Thm 3.6.4, for given $\varepsilon_n \downarrow 0$, the laws of $\{(\mu_t^{\varepsilon_n})_{t \ge 0}\}_{n \in \mathbb{N}}$ are tight on $\mathcal{D}([0, \infty), \mathcal{M}(\mathbb{R}))$ if and only if

- **(J)** (Tightness of evaluations) For each $f \in \mathcal{C}_{\mathrm{c}}^2(\mathbb{R})$, the laws of $\{(\langle \mu_t^{\varepsilon_n}, f \rangle)_{t \ge 0}\}_{n \in \mathbb{N}}$ are tight on $\mathcal{D}([0, \infty), \mathbb{R})$.

To verify tightness of the laws of the real-valued processes $(\langle \mu_t^{\varepsilon_n}, f \rangle)_{t \ge 0}$, we will use the following lemma.

Lemma 2.13 (Tightness criterion). *Let $\xi^n \in \mathcal{D}([0, \infty), \mathbb{R})$ and assume that*

$$\lim_{\delta \downarrow 0} \sum_{i=0}^{\lfloor T/\delta \rfloor} \limsup_{n \to \infty} \mathbb{P}^x \Big[\sup_{t \in [i\delta, (i+1)\delta)} \big| \xi_t^n - \xi_{i\delta}^n \big| \ge \eta \Big] = 0 \quad (\eta > 0, T < \infty). \tag{2.53}$$

Then the laws $\mathbb{P}[\xi^n \in \cdot]$ are tight on $\mathcal{D}([0, \infty), \mathbb{R})$ and each weak limit point is concentrated on $\mathcal{C}([0, \infty), \mathbb{R})$.

Proof. It is well known [6], Thm 15.5, that the conclusion of the lemma is implied by

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{P}^x \Big[\sup_{0 \le s < t \le T : |s - t| \le \delta} \big| \xi_t^n - \xi_s^n \big| \ge \eta \Big] = 0 \quad (\eta > 0, T < \infty). \tag{2.54}$$

If $|\xi_t^n - \xi_s^n| \ge \eta$ for some $0 \le s < t \le T$ with $|s - t| \le \delta$, then there must exist $0 \le i \le \lfloor T/\delta \rfloor$ and $i\delta \le s < t < (i+1)\delta$ such that $|\xi_t^n - \xi_s^n| \ge \eta/2$, and hence $\sup_{t \in [i\delta, (i+1)\delta)} |\xi_t^n - \xi_{i\delta}^n| \ge \eta/4$. This shows that (2.53) implies (2.54). $\square$

We will establish tightness for $\{(\mu_t^{\varepsilon_n})_{t \ge 0}\}_{n \in \mathbb{N}}$ by delicate comparisons between biased and unbiased voter models using results from [3] that we now cite.

Lemma 2.14 (Continuity estimate for the unbiased model). *Let $\mathbb{P}^x$ denote the law of the unbiased voter model $(X_t^0)_{t \ge 0}$ started in $X_0^0 = x$ and let*

$$\nu_t^{\varepsilon} := \sum_{i \in \mathbb{Z}} \varepsilon X_{\varepsilon^{-2} t}^0(i) \delta_{\varepsilon i} \quad (\varepsilon > 0, t \ge 0). \tag{2.55}$$

Then for each $f \in \mathcal{C}^2_c(\mathbb{R})$, *there exist* $C < \infty$ *and* $t_0, \varepsilon_0 > 0$ *such that for all* $0 \le t \le t_0$ *and* $0 < \varepsilon \le \varepsilon_0$,

$$\mathbb{P}^x\big[\big|\langle \nu^\varepsilon_t, f\rangle - \langle \nu^\varepsilon_0, f\rangle\big| \ge \delta\big] \le C\delta^{-2}t^{1/4} \quad \big(x \in \{0,1\}^{\mathbb{Z}}, \delta > 0\big). \tag{2.56}$$

Proof. This is proved in Section 2.1 of [3], as a first step towards proving that the laws of the processes in (2.55) are tight. The proof uses the duality between the voter model and coalescing random walks to derive estimates for the mean and variance of $\langle \nu^\varepsilon_t, f\rangle$. Crucially, the bounds are independent of the initial state x and only assume the properties (i)–(iii) of the kernel $a(\cdot)$ that we also use. □

Recall that the main result of [3] is that if an unbiased voter model $(X^0_t)_{t\ge0}$ is started in the Heaviside state $x_{\rm hv}$, then $(\nu^\varepsilon_t)_{t\ge0}$ converges to $(1_{\{y\ge\tilde B_t\}}\,\mathrm{d}y)_{t\ge0}$ as $\varepsilon \downarrow 0$, where $\tilde B_t := W_{\sigma^2 t}$ is a Brownian motion with diffusion coefficient σ^2. Indeed, their proof can be extended to more general initial configurations.

Lemma 2.15 (Invariance principle for the unbiased voter model). *Let* $\varepsilon_n \downarrow 0$. *For each* n, *let* $(X^{0,n}_t)_{t\ge0}$ *be an unbiased voter model started in a deterministic initial state and define* $\nu^{\varepsilon_n}_t$ *as in* (2.55) *but with* X^0 *replaced by* $X^{0,n}$. *Assume that* $\nu^{\varepsilon_n}_0$ *converges in the vague topology to* $1_{\{y\ge0\}}\,\mathrm{d}y$ *as* $n \to \infty$. *Then*

$$\mathbb{P}\big[\big(\nu^{\varepsilon_n}_t\big)_{t\ge0} \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(\nu_t)_{t\ge0} \in \cdot\,\big], \tag{2.57}$$

where $\nu_t := 1_{\{y\ge\tilde B_t\}}\,\mathrm{d}y$ *is the Lebesgue measure on a half line whose boundary is given by the Brownian motion* $(\tilde B_t)_{t\ge0}$ *with diffusion coefficient* σ^2.

Proof. When the initial configuration is $x_{\rm hv}$, tightness and convergence in finite dimensional distributions are shown in Sections 2.1 and 2.2 of [3], respectively. To prove tightness, [3] also use Jakubowski's tightness criterion, where Aldous' tightness criterion (see, e.g., [1], Thm 1) is applied to verify **(J)** that is given above Lemma 2.13. A crucial ingredient is the estimate (2.56), which holds for general initial configurations. In view of this, their proof of tightness holds regardless of the initial condition.

The proof of convergence of the finite dimensional distributions is based on a first and second moment calculation using duality and the fact that a collection of dual coalescing random walks converges to a collection of coalescing Brownian motions. For this part of the argument, it suffices if the initial configuration $\nu^{\varepsilon_n}_0$ converges in the vague topology to $1_{\{y\ge0\}}\,\mathrm{d}y$ as $n \to \infty$. □

We now prove tightness by verifying Jakubowski's tightness criterion **(J)**, using Lemmas 2.13, 2.14, and 2.15, and judicious comparisons between biased and unbiased voter models, thereby completing the proof of our main result Theorem 1.1.

Proof of Theorem 1.1. If a is the nearest-neighbor kernel $a(-1) = \frac12 = a(1)$, then the Heaviside state is a trap for the process modulo translations. As in (2.17), let τ^ε_0 denote the trapping time. It has been shown below (2.40) that in this case, the biased voter model observed until τ^ε_0 converges in law to the unbiased voter model observed until τ^0_0, and that τ^0_0 is finite a.s. In view of this, in this case, Theorem 1.1 follows trivially from Theorem 1.2. We assume therefore without loss of generality that a is non-nearest-neighbor.

Convergence of finite dimensional distributions has already been proved in Lemma 2.12, so it suffices to show tightness. As argued at the beginning of this section, by Jakubowski's tightness criterion, it suffices to show that for each $f \in \mathcal{C}^2_c(\mathbb{R})$, the laws of the processes $\{(\langle \mu^{\varepsilon_n}_t, f\rangle)_{t\ge0}\}_{n\in\mathbb{N}}$ are tight on $\mathcal{D}([0,\infty),\mathbb{R})$ along any sequence $\varepsilon_n \downarrow 0$. By linearity, it suffices to consider nonnegative f. We fix

$f \geq 0$ and apply Lemma 2.13 to the real-valued processes $(\langle \mu_t^{\varepsilon_n}, f\rangle)_{t\geq 0}$. We fix $\eta > 0$ and for each n and $s \geq 0$ define

$$\begin{aligned} \tau_s^{n,+} &:= \inf\bigl\{t \geq 0 : \bigl\langle \mu_{s+t}^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_s^{\varepsilon_n}, f\bigr\rangle \geq \eta\bigr\}, \\ \tau_s^{n,-} &:= \inf\bigl\{t \geq 0 : \bigl\langle \mu_{s+t}^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_s^{\varepsilon_n}, f\bigr\rangle \leq -\eta\bigr\}. \end{aligned} \tag{2.58}$$

We will prove the lower and upper bounds

$$\left.\begin{aligned} &\text{(i)}\ \lim_{\delta\downarrow 0} \delta^{-1} \sup_{s\geq 0} \limsup_{n\to\infty} \mathbb{P}\bigl[\tau_s^{n,+} < \delta\bigr] = 0, \\ &\text{(ii)}\ \lim_{\delta\downarrow 0} \delta^{-1} \sup_{s\geq 0} \limsup_{n\to\infty} \mathbb{P}\bigl[\tau_s^{n,-} < \delta\bigr] = 0 \end{aligned}\right\} \quad (\eta > 0), \tag{2.59}$$

which together imply (2.53) and hence tightness for the laws of $(\langle \mu_t^{\varepsilon_n}, f\rangle)_{t\geq 0}$.

To prove (2.59)(i), we note that if $\tau_s^{n,+} < \delta$ and the increment of the process $(\langle \mu_t^{\varepsilon_n}, f\rangle)_{t\geq 0}$ during the remaining time from $s + \tau_s^{n,+}$ to $s + \delta$ is no less than $-\eta/2$, then the increment of the process between s and $s + \delta$ is larger than $\eta/2$. Thus,

$$\mathbb{P}\bigl[\bigl\langle \mu_{s+\delta}^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_s^{\varepsilon_n}, f\bigr\rangle \geq \eta/2\bigr] \geq C_{\delta,n} \mathbb{P}\bigl[\tau_s^{n,+} < \delta\bigr], \tag{2.60}$$

where

$$C_{\delta,n} := \inf_{0\leq t<\delta} \inf_x \mathbb{P}^x\bigl[\bigl\langle \mu_t^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_0^{\varepsilon_n}, f\bigr\rangle \geq -\eta/2\bigr], \tag{2.61}$$

and we have conditioned on $\tau = \tau_s^{n,+}$ and $x = X_{s+\tau}^{\varepsilon_n}$ and used the strong Markov property. We couple the biased voter model started in $X_0^{\varepsilon_n} = x$ to an unbiased voter model started in $X_0^0 = x$ in such a way that $X_t^{\varepsilon_n} \geq X_t^0$ for all $t \geq 0$. Defining $\nu_t^{\varepsilon_n}$ as in (2.55), using that $f \geq 0$, it follows that

$$\mathbb{P}^x\bigl[\bigl\langle \mu_t^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_0^{\varepsilon_n}, f\bigr\rangle \geq -\eta/2\bigr] \geq \mathbb{P}^x\bigl[\bigl\langle \nu_t^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \nu_0^{\varepsilon_n}, f\bigr\rangle \geq -\eta/2\bigr]. \tag{2.62}$$

Using the symmetry between zeros and ones in the unbiased voter model, we obtain by Lemma 2.14 that $C_{\delta,n} \geq 1 - C\eta^{-2}\delta^{1/4} \geq 1/2$ for all δ small enough and n large enough. Inserting this into (2.60) and using the convergence of the finite dimensional distributions (Lemma 2.12), we find that for δ small enough,

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}\bigl[\tau_s^{n,+} < \delta\bigr] &\leq 2 \limsup_{n\to\infty} \mathbb{P}\bigl[\bigl\langle \mu_{s+\delta}^{\varepsilon_n}, f\bigr\rangle - \bigl\langle \mu_s^{\varepsilon_n}, f\bigr\rangle \geq \eta/2\bigr] \\ &= 2\mathbb{P}\left[\int f(x) 1_{\{x\geq B_{s+\delta}\}}\,\mathrm{d}x - \int f(x) 1_{\{x\geq B_s\}}\,\mathrm{d}x \geq \eta/2\right] \\ &\leq 2\mathbb{P}\left[|B_{s+\delta} - B_s| \geq \frac{1}{2}\eta\|f\|_\infty\right], \end{aligned} \tag{2.63}$$

where $B_t = W_{\sigma^2 t} - \frac{1}{2}\sigma^2 t$ is a Brownian motion with drift $-\frac{1}{2}\sigma^2$ and diffusion constant σ^2. It is easy to see the right-hand side is $o(\delta)$, uniformly in s, proving (2.59)(i).

The argument for (2.59)(ii) is similar, but not quite the same. In this case, we couple biased and unbiased voter models started in the same initial state at time s to bound

$$\mathbb{P}\bigl[\tau_s^{n,-} < \delta\bigr] \leq \mathbb{P}\bigl[\sigma_s^{n,-} < \delta\bigr], \tag{2.64}$$

where

$$\sigma_s^{n,-} := \inf\{t \geq 0 : \langle \nu_{s+t}^{\varepsilon_n}, f\rangle - \langle \nu_s^{\varepsilon_n}, f\rangle \leq -\eta\} \tag{2.65}$$

and we use that $f \geq 0$. Arguing as in (2.60) and (2.61), applying Lemma 2.14 directly without the need of the coupling in (2.62), allows us to estimate, for δ small enough,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}[\sigma_s^{n,-} < \delta] &\leq 2 \limsup_{n\to\infty} \mathbb{P}[\langle \nu_{s+\delta}^{\varepsilon_n}, f\rangle - \langle \nu_s^{\varepsilon_n}, f\rangle \geq \eta/2] \\ &= 2\mathbb{P}\Big[\int f(x) 1_{\{x \geq \tilde{B}_{s+\delta}\}}\,\mathrm{d}x - \int f(x) 1_{\{x \geq \tilde{B}_s\}}\,\mathrm{d}x \geq \eta/2\Big] \\ &\leq 2\mathbb{P}\Big[|\tilde{B}_{s+\delta} - \tilde{B}_s| \geq \frac{1}{2}\eta \|f\|_\infty\Big],\end{aligned} \tag{2.66}$$

where we have used Lemma 2.15 instead of Lemma 2.12 and $(\tilde{B}_t)_{t\geq s}$ is a Brownian motion with zero drift and diffusion constant σ^2, started at time s in $\tilde{B}_s = B_s$. Again, the right-hand side is $o(\delta)$, which together with (2.64) proves (2.59)(ii). □

Appendix

A.1. Locally uniform convergence

For any metrizable space E, we let $\mathcal{D}([0,\infty), E)$ denote the space of càdlàg functions (i.e., right-continuous functions with left limits) $w : [0,\infty) \to E$, equipped with the Skorohod topology [6,9], and we let $\mathcal{C}([0,\infty), E)$ denote the subspace of continuous functions. It is well-known that $\mathcal{D}([0,\infty), E)$ is Polish if E is [9], Thm 3.5.6. Moreover, a sequence $w_n \in \mathcal{D}([0,\infty), E)$ converges to a limit $w \in \mathcal{C}([0,\infty), E)$ if and only if $w_n \to w$ locally uniformly on compact sets [9], Lemma 3.10.1. We recall the following well-known lemma.

Lemma A.1 (Convergence criterion). *Let E be a metrizable space and let d be any metric generating the topology on E. Let $w_n, w : [0,\infty) \to E$ be functions and assume that w is continuous. Then $w_n \to w$ locally uniformly if and only if $w_n(t_n) \to w(t)$ for all $t_n, t \geq 0$ such that $t_n \to t$.*

It is not hard to see that locally uniform convergence of functions implies locally uniform convergence of their compositions and inverses. Moreover, for monotone functions, pointwise convergence is equivalent to locally uniform convergence.

Lemma A.2 (Convergence of composed functions). *Let E be a metrizable space, let $\lambda_n, \lambda : [0,\infty) \to [0,\infty)$ be nondecreasing, and let $w_n, w : [0,\infty) \to E$. Assume moreover that λ, w are continuous. Then $\lambda_n \to \lambda$ locally uniformly and $w_n \to w$ locally uniformly imply that $w_n \circ \lambda_n \to w \circ \lambda$ locally uniformly.*

Proof. By Lemma A.1, $t_n \to t$ implies $\lambda_n(t_n) \to \lambda(t)$ and hence $w_n(\lambda(t_n)) \to w(\lambda(t))$. Since this holds for general $t_n \to t$, the claim now follows from Lemma A.1. □

Lemma A.3 (Convergence of nondecreasing functions). *Let $\lambda_n, \lambda : [0,\infty) \to [0,\infty)$ be nondecreasing, and assume that λ is continuous. Let $D \subset [0,\infty)$ be dense. Then $\lambda_n \to \lambda$ locally uniformly if and only if for all $t \in D$ there exist $t_n \geq 0$ such that $t_n \to t$ and $\lambda_n(t_n) \to \lambda(t)$.*

Proof. The necessity of the condition is clear. To prove the sufficiency, by Lemma A.1 it suffices to show that $t_n \to t$ implies $\lambda_n(t_n) \to \lambda(t)$. Fix $t^\pm \in D$ with $t^- < t < t^+$ and choose $t_n^\pm \geq 0$ such that $t_n^\pm \to t^\pm$ and $\lambda_n(t_n^\pm) \to \lambda(t^\pm)$. Then $t_n^- < t_n < t_n^+$ for n sufficiently large, and hence, since the λ_n are nondecreasing, $\lambda_n(t_n^-) \leq \lambda_n(t_n) \leq \lambda_n(t_n^+)$ for n sufficiently large. It follows that $\lambda(t_-) \leq \liminf_{n\to\infty} \lambda_n(t_n)$ and $\limsup_{n\to\infty} \lambda_n(t_n) \leq \lambda(t_+)$. Using the density of D and the continuity of λ, we conclude that $\lambda_n(t_n) \to \lambda(t)$. $\square$

For any $\lambda \in \mathcal{C}([0,\infty),[0,\infty))$, let $\lambda([0,\infty)) := \{\lambda(t) : t \in [0,\infty)\}$ denote the image of $[0,\infty)$ under λ. If λ is strictly increasing and $\lambda([0,\infty)) = [0,\infty)$, then λ has an inverse λ^{-1}.

Lemma A.4 (Convergence of inverse functions). *Let $\lambda_n, \lambda \in \mathcal{C}([0,\infty),[0,\infty))$ be strictly increasing with $\lambda_n([0,\infty)) = [0,\infty)$ and $\lambda([0,\infty)) = [0,\infty)$. Then $\lambda_n \to \lambda$ locally uniformly if and only if $\lambda_n^{-1} \to \lambda^{-1}$ locally uniformly.*

Proof. Let $G := \{(t,\lambda(t)) : t \geq 0\}$ denote the graph of λ and similarly, let G_n denote the graph of λ_n. Then the graph of λ^{-1} is $G^{-1} = \{(\lambda(t),t) : t \geq 0\}$ and similarly for the graph G_n^{-1} of λ_n^{-1}. Lemma A.3 tells us that $\lambda_n \to \lambda$ locally uniformly if and only if for all $(t,s) \in G$ there exist $(t_n,s_n) \in G_n$ such that $(t_n,s_n) \to (t,s)$. Clearly, this holds if and only if G_n^{-1} and G^{-1} satisfy the same condition, which is equivalent to $\lambda_n^{-1} \to \lambda^{-1}$ locally uniformly. $\square$

Let X_n be random variables taking values in a Polish space E, and let $x \in E$. Then it is not hard to see that the following statements are equivalent

1. $\mathbb{P}[X_n \in \cdot] \underset{n\to\infty}{\Longrightarrow} \delta_x$,
2. $\mathbb{P}[X_n \notin A] \underset{n\to\infty}{\longrightarrow} 0$ for all $A \in \mathcal{N}_x$,

where $\Rightarrow$ denotes weak convergence of probability measures and $\mathcal{N}_x$ is a fundamental system of neighborhoods of x. If these conditions are fulfilled, then we say that the X_n converge to x *in probability* and denote this as $X_n \overset{\mathrm{P}}{\to} x$. In particular, let E be a Polish space and d a metric generating the topology on E, let W_n be random variables with values in $\mathcal{D}_E[0,\infty)$, and let $w \in \mathcal{C}_E[0,\infty)$. Then $W_n \overset{\mathrm{P}}{\to} w$ with respect to the Skorohod topology if and only if

$$\sup_{0\leq t\leq T} d\big(W_n(t), w(t)\big) \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} 0 \quad (T < \infty). \tag{A.1}$$

Because we will need these in our proofs, for completeness, we provide proofs for two additional simple lemmas which lift Lemmas A.2 and A.3 to convergence in law and in probability, respectively.

Lemma A.5 (Convergence of time-changed processes). *Let $Y = (Y_t)_{t\geq 0}$ and $Y^n = (Y_t^n)_{t\geq 0}$ be stochastic processes with càdlàg sample paths, taking values in a Polish space E. Let $S = (S_t)_{t\geq 0}$ and $S^n = (S_t^n)_{t\geq 0}$ be real-valued stochastic processes whose sample paths are càdlàg, nondecreasing, and satisfy $S_0 = 0$ resp. $S_0^n = 0$. Assume moreover that $Y = (Y_t)_{t\geq 0}$ and $S = (S_t)_{t\geq 0}$ have continuous sample paths, and that*

$$\mathbb{P}\big[\big(Y_t^n, S_t^n\big)_{t\geq 0} \in \cdot\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(Y_t, S_t)_{t\geq 0} \in \cdot\big], \tag{A.2}$$

where $\Rightarrow$ denotes weak convergence with respect to the Skorohod topology. Then

$$\mathbb{P}\big[\big(Y^n_{S_t^n}\big)_{t\geq 0} \in \cdot\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(Y_{S_t})_{t\geq 0} \in \cdot\big]. \tag{A.3}$$

Proof. By the Skorohod representation theorem [6], Thm 6.7, we can couple our random variables such that $(Y^n_t, S^n_t)_{t\geq 0}$ converges a.s. to $(Y_t, S_t)_{t\geq 0}$ with respect to the Skorohod topology. Now Lemma A.2 implies that $(Y^n_{S^n_t})_{t\geq 0}$ converges a.s. to $(Y_{S_t})_{t\geq 0}$ w.r.t. the same topology, and hence (A.3) follows. $\square$

Lemma A.6 (Convergence of nondecreasing functions). *Let $S^n = (S^n_t)_{t\geq 0}$ be real-valued stochastic processes whose sample paths are càdlàg, nondecreasing, and satisfy $S_0 = 0$ resp. $S^n_0 = 0$. Let $\lambda : [0,\infty) \to [0,\infty)$ be continuous. Then the following statements are equivalent*:

$$\text{(i)} \quad \sup_{0\leq t\leq T} \big|S^n_t - \lambda_t\big| \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} 0 \quad (T<\infty), \qquad \text{(ii)} \quad S^n_t \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} \lambda_t \quad (t\geq 0).$$

Proof. The implication (i)⇒(ii) is trivial. To prove the converse, let $\{t_k : k\in\mathbb{N}\}$ be countable and dense. Then

$$\sup_{0\leq k\leq m} \big|S^n_{t_k} - \lambda_{t_k}\big| \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} 0 \quad (m<\infty), \tag{A.4}$$

which says that the process $k\mapsto S^n_{t_k}$ converges in probability to $k\mapsto \lambda_{t_k}$ with respect to the product topology on $\mathbb{R}^{\mathbb{N}}$. By the Skorohod representation theorem [6], Thm 6.7, we can couple our random variables such that

$$S^n_{t_k} \underset{n\to\infty}{\longrightarrow} \lambda_{t_k} \quad \text{a.s. } (k\in\mathbb{N}). \tag{A.5}$$

By Lemma A.3, it follows that $\sup_{0\leq t\leq T} |S^n_t - \lambda_t|$ converges a.s. to zero for all $T<\infty$, which implies (i). $\square$

A.2. A weak law of large numbers

In this subsection, we prove two simple versions of the weak law of large numbers. Lemma A.8 below is used in the proof of Theorem 1.2. The following lemma would be completely standard if the law of $(V_{n,i})_{i\geq 1}$ would not depend on n.

Lemma A.7 (A weak law of large numbers). *For each $n\geq 1$, let $(V_{n,i})_{i\geq 1}$ be i.i.d. nonnegative random variables, and let $m_n \geq 1$ be integers such that $\lim_{n\to\infty} m_n = \infty$. Assume that*

$$\sup_{n\geq 1} \mathbb{E}\big[|V_{n,1}|\big] < \infty \quad \text{and} \quad \mathbb{E}\big[|V_{n,1}|; |V_{n,1}| > t m_n\big] \underset{n\to\infty}{\longrightarrow} 0 \quad (t>0). \tag{A.6}$$

Then

$$\mathbb{E}[V_{n,1}] - \frac{1}{m_n}\sum_{i=1}^{m_n} V_{n,i} \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} 0, \tag{A.7}$$

where $\overset{\mathrm{P}}{\to}$ denotes convergence in probability.

Proof. Define truncated random variables by $\overline{V}_{n,i} := V_{n,i} 1_{\{|V_{n,i}|\leq m_n\}}$. Then (A.6) implies that

$$\mathbb{P}\left[\sum_{i=1}^{m_n} \overline{V}_{n,i} \neq \sum_{i=1}^{m_n} V_{n,i}\right] \leq m_n \mathbb{P}\big[|V_{n,1}| > m_n\big] \leq \mathbb{E}\big[|V_{n,1}|; |V_{n,1}| > m_n\big] \underset{n\to\infty}{\longrightarrow} 0. \tag{A.8}$$

Since (A.6) moreover implies that

$$\left|\mathbb{E}[\overline{V}_{n,1}] - \mathbb{E}[V_{n,1}]\right| \leq \mathbb{E}\big[|V_{n,1}|;\, |V_{n,1}| > m_n\big] \underset{n\to\infty}{\longrightarrow} 0, \tag{A.9}$$

it suffices to prove the statement with $V_{n,i}$ replaced by $\overline{V}_{n,i}$. For any $\delta > 0$, Chebyshev's inequality gives

$$P\left[\left|\frac{1}{m_n}\sum_{i=1}^{m_n}\overline{V}_{n,i} - \mathbb{E}[\overline{V}_{n,1}]\right| > \delta\right] \leq \delta^{-2}\frac{1}{m_n}\mathrm{Var}(\overline{V}_{n,1}). \tag{A.10}$$

Using the layer cake representation, we estimate

$$\mathrm{Var}(\overline{V}_{n,1}) \leq \mathbb{E}\big[\overline{V}_{n,1}^{2}\big] \leq 2\int_0^{m_n} x\mathbb{P}\big[|V_{n,1}| > x\big]\,\mathrm{d}x \leq 2\int_0^{m_n} \mathbb{E}\big[|V_{n,1}|;\, |V_{n,1}| > x\big]\,\mathrm{d}x. \tag{A.11}$$

It follows that the right-hand side of (A.10) can be estimated by

$$2\delta^{-2}\int_0^1 \mathbb{E}\big[|V_{n,1}|;\, |V_{n,1}| > tm_n\big]\,\mathrm{d}t, \tag{A.12}$$

which tends to zero by (A.6), using dominated convergence. □

Lemma A.8 (Functional law of large numbers). *For each $n \geq 1$, let $(V_{n,i})_{i\geq 1}$ be i.i.d. nonnegative random variables, and let $\varepsilon_n > 0$ be constants such that $\lim_{n\to\infty}\varepsilon_n = 0$. Assume that*

$$\lim_{n\to\infty}\mathbb{E}[V_{n,1}] = c < \infty \quad \text{and} \quad \mathbb{E}[V_{n,1};\, V_{n,1} > t/\varepsilon_n] \underset{n\to\infty}{\longrightarrow} 0 \quad (t > 0). \tag{A.13}$$

Define $f_n : [0,\infty) \to [0,\infty)$ by

$$f_n(t) := \varepsilon_n \sum_{i=1}^{\lfloor \varepsilon_n^{-1} t\rfloor} V_{n,i} \quad (t \geq 0). \tag{A.14}$$

Then

$$\sup_{0\leq t\leq T}\left|ct - f_n(t)\right| \underset{n\to\infty}{\overset{\mathrm{P}}{\longrightarrow}} 0 \quad (T > 0), \tag{A.15}$$

where $\overset{\mathrm{P}}{\to}$ denotes convergence in probability.

Proof. Since the $V_{n,i}$ are nonnegative, the condition $\lim_{n\to\infty}\mathbb{E}[V_{n,1}] = c < \infty$ implies that $\sup_{n\geq n_0}\mathbb{E}[V_{n,1}] < \infty$ for n_0 sufficiently large. Applying Lemma A.7 to $m_n = \lfloor \varepsilon_n^{-1} t\rfloor$, we see that $f_n(t)$ converges in probability to ct for each fixed $t > 0$. The claim now follows from Lemma A.6. □

A.3. Uniform ergodicity

The following lemma, which we apply in Section 2.4, gives sufficient conditions for the speed of convergence to equilibrium to be uniform for a sequence of continuous-time Markov chains.

Lemma A.9 (Uniform ergodicity). *Let S be a countable set and for each $n \in \mathbb{N} \cup \{\infty\}$, let $X^n = (X^n_t)_{t\geq 0}$ be a positive recurrent, irreducible continuous-time Markov chain with state space S and invariant law π_n. Assume that as $n \to \infty$, the jump rates of X^n converge pointwise to the jump rates of X^∞, and the invariant laws π_n converge weakly to π_∞. Then, for each $x \in S$,*

$$\sup_{n\in\mathbb{N}\cup\{\infty\}} \big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \pi_n\big\| \underset{t\to\infty}{\longrightarrow} 0, \tag{A.16}$$

where $\|\cdot\|$ denotes the total variation norm.

Proof. Fix $z \in S$. Let $(X^n_t)_{t\geq 0}$ and $(\tilde{X}^n_t)_{t\geq 0}$ be independent processes with the same jump rates and let $\tau^n_{(z,z)} := \inf\{t \geq 0 : X_t = z = \tilde{X}_t\}$. Using Doeblin's coupling, we can couple two processes by declaring them to be equal after $\tau^n_{(z,z)}$. Therefore, we see that

$$\big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \mathbb{P}^y\big[X^n_t \in \cdot\big]\big\| \leq \mathbb{P}^{(x,y)}\big[t < \tau^n_{(z,z)}\big] \quad (x, y \in S, t \geq 0), \tag{A.17}$$

and hence

$$\begin{aligned} \big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \pi_n\big\| &= \Big\|\sum_{y\in S} \pi_n(y)\big(\mathbb{P}^x\big[X^n_t \in \cdot\big] - \mathbb{P}^y\big[X^n_t \in \cdot\big]\big)\Big\| \\ &\leq \sum_{y\in S} \pi_n(y)\mathbb{P}^{(x,y)}\big[t < \tau^n_{(z,z)}\big] \quad (x \in S, t \geq 0). \end{aligned} \tag{A.18}$$

Since the jump rates converge, the probability $\mathbb{P}^{(x,y)}[t < \tau^n_{(z,z)}]$ converges pointwise as $n \to \infty$ for each $y \in S$. Using also that $\pi_n \Rightarrow \pi_\infty$, which implies that the measures π_n are tight, this is easily seen to imply that the right-hand side of (A.18) converges and hence

$$\limsup_{n\to\infty} \big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \pi_n\big\| \leq \sum_{y\in S} \pi_\infty(y)\mathbb{P}^{(x,y)}\big[t < \tau^\infty_{(z,z)}\big] \quad (x \in S, t \geq 0). \tag{A.19}$$

The joint process $(X^\infty_t, \tilde{X}^\infty_t)_{t\geq 0}$ is irreducible and has an invariant law $\pi_\infty \otimes \pi_\infty$, which implies positive recurrence. In view of this, the right-hand side of (A.19) converges to zero as $t \to \infty$ for each fixed $x \in S$. Since X^n is positive recurrent and hence ergodic for each $n \in \mathbb{N}$ and since the total variation distance to the invariant measure is a nonincreasing function of time,

$$\begin{aligned} \limsup_{t\to\infty} \sup_n \big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \pi_n\big\| &\leq \limsup_{t\to\infty} \sup_{n\geq N} \big\|\mathbb{P}^x\big[X^n_t \in \cdot\big] - \pi_n\big\| \\ &\leq \sup_{n\geq N} \big\|\mathbb{P}^x\big[X^n_T \in \cdot\big] - \pi_n\big\| \end{aligned} \tag{A.20}$$

for each $N, T < \infty$, where in view of (A.19) the right-hand side can be made arbitrary small by choosing N and T large enough. □

Acknowledgements

R. Sun is supported by NUS grant R-146-000-253-114. J.M. Swart is supported by grant 19-07140S of the Czech Science Foundation (GA CR). J. Yu is supported by a grant from NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai.

References

[1] Aldous, D. (1978). Stopping times and tightness. *Ann. Probab.* **6** 335–340. MR0474446 https://doi.org/10.1214/aop/1176995579

[2] Asmussen, S. (2003). *Applied Probability and Queues*, 2nd ed. *Applications of Mathematics* (*New York*): *Stochastic Modelling and Applied Probability* **51**. New York: Springer. MR1978607

[3] Athreya, S.R. and Sun, R. (2011). One-dimensional voter model interface revisited. *Electron. Commun. Probab.* **16** 792–800. MR2868600 https://doi.org/10.1214/ECP.v16-1688

[4] Belhaouari, S., Mountford, T., Sun, R. and Valle, G. (2006). Convergence results and sharp estimates for the voter model interfaces. *Electron. J. Probab.* **11** 768–801. MR2242663 https://doi.org/10.1214/EJP.v11-349

[5] Belhaouari, S., Mountford, T. and Valle, G. (2007). Tightness for the interfaces of one-dimensional voter models. *Proc. Lond. Math. Soc.* (3) **94** 421–442. MR2308233 https://doi.org/10.1112/plms/pdl016

[6] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. New York: Wiley. MR1700749 https://doi.org/10.1002/9780470316962

[7] Cox, J.T. and Durrett, R. (1995). Hybrid zones and voter model interfaces. *Bernoulli* **1** 343–370. MR1369166 https://doi.org/10.2307/3318488

[8] Dawson, D.A. (1993). Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI*—1991. *Lecture Notes in Math.* **1541** 1–260. Berlin: Springer. MR1242575 https://doi.org/10.1007/BFb0084190

[9] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838085 https://doi.org/10.1002/9780470316658

[10] Fontes, L.R.G., Isopi, M., Newman, C.M. and Ravishankar, K. (2004). The Brownian web: Characterization and convergence. *Ann. Probab.* **32** 2857–2883. MR2094432 https://doi.org/10.1214/009117904000000568

[11] Newman, C.M., Ravishankar, K. and Sun, R. (2005). Convergence of coalescing nonsimple random walks to the Brownian web. *Electron. J. Probab.* **10** 21–60. MR2120239 https://doi.org/10.1214/EJP.v10-235

[12] Sun, R. and Swart, J.M. (2008). The Brownian net. *Ann. Probab.* **36** 1153–1208. MR2408586 https://doi.org/10.1214/07-AOP357

[13] Sun, R., Swart, J.M. and Yu, J. (2019). Equilibrium interfaces of biased voter models. *Ann. Appl. Probab.* **29** 2556–2593. MR3984257 https://doi.org/10.1214/19-AAP1461

Received March 2020 and revised July 2020