Spatial Moran Models
II. Tumor growth and progression

Richard Durrett[1*], Jasmine Foo[2†], and Kevin Leder [3†]

1. Dept of Mathematics, Duke U., Durham NC

2. Dept of Mathematics and 3. Industrial and Systems Engineering,

U. of Minnesota, Minneapolis, MN

August 9, 2013

Abstract

We study the accumulation of mutations in a spatial Moran model on a torus in $\mathbb{Z}^d$ in which each cell gives birth at a rate equal to its fitness and replaces a neighbor at random with its offspring. Cells of type k have relative fitness $(1+s)^k$ and mutate to type $k+1$ at rate u_{k+1}. When restricted to two cell types and no mutations, this model reduces to the biased voter model. We give a new result for the biased voter model that identifies the asymptotic behavior of the speed of propagation in the Bramson-Griffeath shape theorem, when $s \to 0$. Motivated by our results on the spatial Moran model we introduce a simplified model and in the context of this model we study σ_k, the time of birth of the first type k whose family line does not die out, and the growth of the number of type k cells, $Z_k(t)$. This investigation is a first step in understanding the spatial structure of the genetic heterogeneity of solid tumors.

1 Introduction

Cancer initiation and progression are driven by the accumulation of mutations in cell populations. These mutations can confer changes to cellular reproduction rates and enable rapid growth and evolution of tumors. Understanding the dynamics of mutation accumulation in fixed or exponentially growing populations contributes to a better understanding of when and how cancers initiate, as well as the genetic diversity of tumors.

*Partially supported by NIH grant 5R01GM096190

†Partially supported by NSF grant DMS-1224362

There has been a substantial amount of previous modeling effort devoted to the study of temporal dynamics of mutation accumulation in cancer initiation and progression. These works have primarily been restricted to the setting of homogeneously mixed populations, and fallen into the following three categories: multi-type Moran models with a homogeneously mixing population of either (i) constant or (ii) exponentially growing size, and (iii) multi-type branching processes. We refer the reader to Part I of this work by Durrett and Moseley (2012) for a discussion of the literature in category (i). Systems of type (ii) have been studied by Beerenwinkel et al. (2007) and Durrett and Mayberry (2007), who have shown that there are traveling waves of selective sweeps. In category (iii) Durrett and Moseley (2009) and Bozic et al. (2010) studied systems in which individuals of type k always have fitness $(1+s)^k$. Durrett, Foo, Leder, Mayberry and Michor (2011a,2011b) generalized the results to the situation where fitness advances are random, and studied heterogeneity.

Although these previous studies have led to a better understanding of the temporal process of carcinogenesis in well-mixed populations, recent studies have revealed that cancer evolution is strongly impacted by spatial dynamics within the tissues in which cancers arise (e.g. Chai et. al. 2009, Martens et. al. 2011). In particular, carcinogenesis is driven by a spatiotemporal process involving the accumulation and spatial expansion of one or more patches of mutated cells within a homeostatic compartment; the growth rate and geometry of these patches impact the probability of subsequent mutations and timing of cancer initiation. The main goal of this work is to understand these dynamics by considering a simple two-step mutational pathway; to this end we will consider (iv) spatial Moran models of constant size.

Historically, the first spatial model of tumor growth is the one of Williams and Bjerknes (1972). In this model, there are two cell types: 0 (healthy) and 1 (tumor), with fitnesses 1 and $\lambda > 1$. The cell at x gives birth at a rate equal to its fitness to an offspring that replaces the cell at one of the $2d$ nearest neighbors chosen at random. Bramson and Griffeath (1980,1981) proved the first rigorous results about the asymptotic behavior of this model, which is also called the biased voter model. In particular they proved a "shape theorem" for the asymptotic behavior of the process which is stated and used below. In related work, Martens and Hallatschek used simulation to study the dynamics of mutation accumulation and population adaptation on a discrete time hexagonal lattice model with nearest neighbor replacement upon reproduction in one and two dimensions. Here, the replacement rule in each generation was tuned so that adaptation waves travel at a desired speed that matches classical estimates of Fisher wave speeds. In addition, an additional probability of death was imposed at the birth of each mutant in order to calibrate the dynamics to a desired survival probability for each new mutant clone. Under this model, using small fitness advantages s, the authors studied the speed of rate of change of the mean fitness advantage per generation and found that this speed saturates once the domain size exceeds a characteristic interference length. In a follow-up work, this model was applied to study the impact of clonal interference on the waiting time for

cancer initiation. Although the authors considered a similar question to the current work, we note that in these works the microscopic behaviors are artificially tuned to match desired macroscopic phenomena; then, the tuned model is simulated to characterize other macroscopic behaviors. In contrast, here we begin by assuming only the microscopic dynamics of reproduction, death and mutation on the lattice, making no prior assumptions about the macroscopic dynamics. We then analyze this process to characterize the emergent macroscopic behaviors and their dependence on microscopic parameters.

In Part I, Durrett and Moseley (2012) considered a spatial Moran model in which there are three types of cells (0, 1, and 2) all with fitness 1, and type i cells mutate to type $i+1$ cells at rate u_i. Under these assumptions, limit theorems for τ_2 the time of the birth of the first type 2 were proved. These results were an extension of (Komarova, 2007), in which the rate of producing two-hit neutral mutants in a 1-dimensional process was studied and compared to the well-mixed setting. In the present paper we focus on the case in which new mutations have a selective advantage over the previous ones. Throughout the paper we will use the following notation for the asymptotic behavior of positive functions.

$$\begin{aligned} f(t) \sim g(t) &\quad \text{if } f(t)/g(t) \to 1 \text{ as } t \to \infty \\ f(t) = o(g(t)) \text{ or } \ll g(t) &\quad \text{if } f(t)/g(t) \to 0 \text{ as } t \to \infty \\ f(t) \gg g(t) &\quad \text{if } f(t)/g(t) \to \infty \text{ as } t \to \infty \\ f(t) = O(g(t)) &\quad \text{if } f(t) \le Cg(t) \text{ for all } t \\ f(t) = \Theta(g(t)) &\quad \text{if } cg(t) \le |f(t)| \le Cg(t) \text{ for all } t \end{aligned}$$

The outline of the remainder of the paper is as follows. In Section 2 we introduce the spatial Moran model and state our main two results concerning this process. Based on these two results we introduce our simplified model with deterministic growth. In Section 3 we state our primary results regarding the timing of successful mutations in the simplified model. Section 4 is concerned with an application of this material to colon cancer, and section 5 reviews some basic results regarding the biased voter model. Section 6 is devoted to proofs of useful facts about the biased voter model. Theorems 2 is proved in Section 7, Theorems 3 and 4 in Section 8, and Theorem 5 in Section 9. The proof of Theorem 1 is hidden away in Section 10 because it relies on different techniques, and follows easily from results of Durrett and Zähle (1997).

2 A spatial Moran model of tumor growth

Although we are mainly interested in studying processes on the torus $[-L, L]^d \cap \mathbb{Z}^d$, we first consider a spatial model of tumor growth on a lattice in $\mathbb{Z}^d$. Type i cells have fitness $(1+s)^i$. Type i cells mutate to become type $i+1$ cells at rate u_i. Let

$\phi_t(z)$ be the fitness of the cell at z at time t. If we consider two nearest neighbors x and y then the value at x will replace the one at y at rate $\phi_t(x)$ and y will replace x at rate $\phi_t(y)$. The spatial Moran model we have just described can be viewed as an extension of the biased voter model. In particular, the biased voter model is the spatial Moran model restricted to type 0 and type 1 cells, without any mutations. To study our model, it will be useful to describe some properties of the biased voter model.

2.1 Growth from a single type 1

Consider a biased voter model with two types 0 and 1 with fitnesses 1 and $\lambda > 1$. Then at each boundary edge connecting a 1 with a 0, the first event will be the 0 changing to 1 with probability $p = \lambda/(\lambda+1)$ or the 1 changing to 0 with probability $1-p = 1/(\lambda+1)$. Thus if $\xi_t = \{x : \phi_t(x) = \lambda\}$ then while $\xi_t \neq \emptyset$, the size of the set, $|\xi_t|$, is a biased random walk which makes jumps at rate equal to $1+\lambda$ times the number of boundary edges: $|\partial\xi_t| = |\{x \sim y : x \in \xi_t, y \notin \xi_t\}|$. Here we have used $x \sim y$ to indicate that x and y are neighbors even though this conflicts with our previously announced notation.

Elementary random walk results imply that if we start with one type 1 and let $T_k = \inf\{t : |\xi_t| = k\}$ then using a subscript of 1 to indicate starting from one individual with fitness λ

$$P_1(T_0 = \infty) = 1 - \frac{1-p}{p} = 1 - \frac{1}{\lambda} = \frac{\lambda-1}{\lambda} = \frac{s}{1+s}$$

if $\lambda = 1+s$. Maruyama (1970, 1974) was the first to notice that the fixation probability is not changed by considering a spatial model, but this fact has been rediscovered by others, see Lieberman, Hauert, and Nowak (2005). Even in cancer, selective advantages s are small, so we will use the approximation

$$P_1(T_0 = \infty) \approx s. \tag{1}$$ notdie

Let ξ_t^0 be the set of sites occupied by individuals of type 1 at time t when initially there is a single 1 at the origin at time 0. Bramson and Griffeath (1980, 1981) showed that when ξ_t^0 does not die out, it grows linearly and has an asymptotic shape D. That is, for any $\epsilon > 0$, there is a t_ϵ (which depends on the outcome ω) so that on $\{T_0 = \infty\}$ we have

$$(1-\epsilon)tD \cap \mathbb{Z}^d \subset \xi_t \subset (1+\epsilon)tD \quad \text{for } t \geq t_\epsilon(\omega).$$

D is convex and has the same symmetries as $\mathbb{Z}^d$.

Let e_1 be the first unit vector and define the growth rate $c_d(s)$ so that the intersection of D with the x axis is $[-c_d(s)e_1, c_d(s)e_1]$. It is easy to compute $c_1(s)$. If $\xi_t \neq \emptyset$ then $\xi_t = [l_t, r_t]$. The right edge r_t increases by 1 at rate λ and decreases by 1 at rate 1, so $r_t/t \to \lambda - 1 = s$, i.e., $c_1(s) = s$. The proof of Bramson and Griffeath implies that $c_d(s) \geq b_d s$ where b_d is a positive constant. By using techniques of Durrett and Zähle (2007), we can find the order of magnitude of the speed.

DZ **Theorem 1.** *As $s \to 0$ we have*

$$c_d(s) \sim \begin{cases} s & d = 1 \\ \sqrt{4\pi s/\log(1/s)} & d = 2 \\ \sqrt{4\beta_d s} & d \geq 3, \end{cases}$$

where β_d is the probability that two d dimensional simple random walks started at 0 and $e_1 = (1, 0, \ldots 0)$ never hit.

The proof of this result is found in Section 10.

2.2 Time of the first successful type 1 mutant

Let σ_i be the time of birth of the first successful type i, i.e., one whose family line does not die out. Since we are considering the biased voter model on a finite set, $[-L, L]^d \cap \mathbb{Z}^d$, the first issue to consider is: what does it mean for a mutation to not die out? When $\lambda = 1 + s$ formulas (20) and (21) (for the biased voter model on $\mathbb{Z}^d$) imply that

$$P_1(T_k < T_0) = \frac{1 - (1+s)^{-1}}{1 - (1+s)^{-k}} \qquad P_k(T_0 < \infty) = \lambda^{-k}$$

When $s \to 0$ and $k \sim C/s$

$$P_1(T_k < T_0) \to \frac{s}{1 - e^{-C}} \qquad P_k(T_0 < \infty) \to e^{-C},$$

so an appropriate finite size version of survival is reaching size C/s, where C is large. To have success probability $\approx s$, the value on $\mathbb{Z}^d$, we will let $C_s \to \infty$ slowly as $s \to 0$.

Since mutations occur at rate Nu_1 and are successful with probability s, it is almost obvious that:

sig1lim **Theorem 2.** *If $s, u_1 \to 0$ then $P(\sigma_1 > t/Nu_1 s) \to e^{-t}$.*

While this is intutive, we have not been able to prove this without an additional technical assumption. Results in Section 6 will show that unsuccessful type 1 mutations will typically die out by a time of order

$$\ell(s) = \begin{cases} s^{-2} & d = 1 \\ s^{-1}\log(1/s) & d = 2 \\ s^{-1} & d \geq 3 \end{cases} \tag{2}$$ tofs

Since particles in the dual process move like random walks, they can move a distance $O(\ell(s)^{1/2})$ in time $\ell(s)$. Thus, an unsuccessful type 1 mutation can spread over a space-time volume of $O(\ell(s)^{(d+2)/2})$. To be able to easily estimate the dependence between different trials, we will assume

$$(A0) \qquad (1/u_1) \gg \ell(s)^{(d+2)/2} \tag{3}$$ A0

This assumption will be in force throughout the paper. For a typical value of $s = 0.01$ (see the discussion in Section 4 for more on this) the condition is satisfied when $u_1 < 10^{-6}$. Unfortunately this is not satisfied in our example in Section 4.2, we have that $u_1 \approx 10^{-5}$.

2.3 A simplified model with deterministic growth

Theorems 1, 2 and the Bramson-Griffeath shape theorem suggests a simplification to our model.

- When a type I mutation occurs, it can either be a successful type 1 (which corresponds intuitively to its family line not dying out and occurs with probability s), or it can be an unsuccessful type 1 with probability $1 - s$.
- Successful mutations initiate expanding clones that are growing balls (in the usual ℓ_2 norm) whose radius at time t is $c_d(s)t$. Note that here we are referring to the balls as solid objects in $\mathbb{R}^d$, so one must intersect the solid ball with $\mathbb{Z}^d$.
- Unsuccessful type 1 mutations give rise to a copy of the biased voter model conditioned to die out.

Chatterjee and Durrett (2011) took the same approach in their study of Aldous' gossip processes, in which information can spread by long range jumps in addition to a nearest neighbor process that is first passage percolation. Rather than wrestle with the details of the estimating the growth from the different seeds, they replaced the random growth process by balloons that grow linearly in radius. In that paper and this one, we do not expect this simplification to significantly alter the qualitative behavior.

Theorems 3–5 will be proved for the simplified model.

If we are in a situation where successful mutations to type $k + 1$ only come from descendants of successful type k mutations, our simplified model is the polynuclear growth (PNG) model of Prähofer and Spohn (2000). In that system, mutations occur at rate 1, and when they land on a point with height k produce a linearly growing disk at height $k + 1$. There is a large literature on the one-dimensional PNG model, which can be studied in great detail thanks to connections to increasing sequences in random permutations and random matrices. See Ferrari and Prähofer (2006), where it is shown that the scaled one dimensional profile of the PNG model converge to solutions of Kardar-Parisi-Zhang (1986) equation. However, very little is known in dimensions $d \geq 2$. and the state of the spatial Moran model after hundreds or thousands of mutations is not relevant to studying cancer.

2.4 Takeover by 1's

Consider now the special case of our simplified model with deterministic growth in which the mutation rates $u_i \equiv u$ and we start with no individuals of type 1, i.e., $\xi_0 = \emptyset$. Since we are interested first in finding the time at which the 1's take over the space, we will ignore the mutations that produce 2's and use ξ_t to indicate the set of sites occupied by 1's. The time until the first successful type 1 mutation will be

$$t_{mut} = \Theta(1/L^d us).$$

The time to takeover the system $t_{fix} = \Theta(L/c_d)$. Setting $t_{mut} = t_{fix}$ and solving we see that if

$$(SF) \qquad L \ll L_c = \left(\frac{c_d}{su}\right)^{1/(d+1)} \tag{4}$$ SF

then we will have *sequential fixation*: mutations will fix faster than they arise, and the times between successive mutations that do not die out will be exponential with mean $1/t_{mut}$.

Let γ_d be the volume of a ball of radius 1 in d dimensions,

$$\gamma_1 = 2, \qquad \gamma_2 = \pi, \qquad \gamma_3 = 4\pi/3 \tag{5}$$ gammadef

A site x will be type 1 at time t if there is a successful type 1 mutation in the space-time cone $\{(y,r) : |y-x| < c_d(t-r)\}$. Such mutations are approximately a Poisson process with rate us so

$$P(x \in \xi_t) \approx 1 - \exp\left(-us\int_0^t \gamma_d (c_d r)^d \, dr\right) \approx 1 - \exp\left(-us\frac{\gamma_d c_d^d t^{d+1}}{d+1}\right) \tag{6}$$ fcover

This quantity will go from a small density ϵ to $1-\epsilon$ at times of order $(1/suc_d^d)^{1/(d+1)} = L_c/c_d$.

To prepare for later developments, we note that the number of successful type 1 mutations by this time will be

$$K_{pos} = \frac{L_c}{c_d} \cdot L^d su = L^d \left(\frac{su}{c_d}\right)^{d/(d+1)} = (L/L_c)^d. \tag{7}$$ Katfix

This observation will be used later in (14) to show that with high probability σ_2 will occur while the density of 1's is small. Because of this, Theorem 2 can be upgraded to show that up to time σ_2 successful type 1 mutations occur at times of a Poisson process with rate $Nu_1 s$.

3 Waiting time for σ_2

sec:sig2

In many cancer types (e.g. breast, colorectal), early mutations in the genetic progression compromise the DNA replication machinery, resulting in a drastically increased mutation rate for subsequent events. Thus it is important to drop the assumption $u_1 = u_2$. In this section we investigate the behavior of σ_2, the waiting time until the first successful type-2 mutant. There are three regimes of behavior, depending on the value of

$$\Gamma = (Nu_1 s)^{d+1}(c_d^d u_2 s)^{-1}, \tag{8}$$

Gammadef

which (as we will see shortly) is related to the number of successful type 1 mutations necessary to create a successful type 2 mutation. Throughout this section, we are taking the limit as $s, u_1, u_2 \to 0$, and proving our results for the simplified model.

3.1 $\Gamma \to 0$

In his Cornell Ph.D. thesis, Stephen Moseley began the study of the asymptotic behavior of the waiting time σ_2. His result is for the regime in which the first successful mutation to type 2 occurs in the first successful type 1 family and before it reaches fixation.

DMth3 **Theorem 3.** *If we assume,*

$$(A1) \qquad \left(\frac{c_d}{u_2 s}\right)^{d/(d+1)} \ll N \ll \frac{(c_d^d u_2 s)^{1/d+1}}{u_1 s} \qquad (A2)$$

and (A3) $u_2 \ll 1/\ell(s)$ *then* $P(\sigma_2 > t/Nu_1 s) \to \exp(-t)$.

Here $\ell(s)$ is the quantity defined in (2). To connect with the title of the subsection note that $(A2)$ is the condition $\Gamma \to 0$. The proof is provided in section 8.

Since the conditions of Theorem 3 are somewhat complicated, we will now explain them intuitively. To see the reason for (A1), we note that under our simplified model, if the successful type 2 mutation occurs before type 1's reach fixation, it will occur $\Theta(t_2)$ units of time after the type 1 mutation when the space-time volume covered by the its descendents

$$\int_0^{t_2} (c_d r)^d \, dr = \Theta(1/u_2 s).$$

That is, $t_2 = \Theta((c_d^d u_2 s)^{-1/(d+1)})$. At that time the radius of the set of 1's

$$c_d t_2 = \Theta((c_d/u_2 s)^{1/(d+1)})$$

For this to fit inside our torus, we need to have

$$(c_d/u_2 s)^{d/(d+1)} \ll L^d = N \tag{9}$$

A1

for the computation of the integral to be valid.

To explain (A2), note that if we let σ_1 be the time of the first successful 1 mutation then by Theorem 2

$$P(\sigma_1 > t/Nu_1s) \approx e^{-t}, \tag{10}$$ sig1lt

so for the result in Theorem 3 to hold we must have $\sigma_2 - \sigma_1 \ll \sigma_1$, which requires

$$t_2 = (c_d^d u_2 s)^{-1/(d+1)} \ll 1/Nu_1 s$$

or rewriting things in terms of N

$$N \ll \frac{(c_d^d u_2 s)^{1/(d+1)}}{u_1 s}. \tag{11}$$ A2

Finally, (A3) is needed to rule out the possibility that the successful mutation to type 2 occurs among the descendants of an unsuccessful type 1 mutation. Lemma 6.2 will show that if ξ_t^0 is the set of 1's in a supercritical biased voter model on $\mathbb{Z}^d$ with $\lambda = 1 + s$ and T_0 is the time at which the process dies out then

$$E\left(\int_0^{T_0} |\xi_t^0|\, dt \middle| T_0 < \infty\right) \le C\ell(s)$$

where $\ell(s)$ was defined in (2). Mutations to type 2 that land on an unsuccessful type 1 family will succeed with a probability between s and $2s$, since when they grow outside the unsuccessful type 1 family they will be competing with type 0's. Since the expected number of type 1 mutations before time t/Nu_1s is $O(1/s)$, the expected number of successful mutations to type 2 that occur in these families is $O(u_2\ell(s)) \to 0$ by (A3).

3.2 $\Gamma \to I \in (0, \infty)$

Ith **Theorem 4.** *If we assume (A1), (A3), and* $\Gamma \to I \in (0, \infty)$ *then*

$$P(\sigma_2 > t/Nu_1s) \to \exp\left(-\int_0^t 1 - \exp\left[-\frac{\gamma_d}{I} \cdot \frac{y^{d+1}}{d+1}\right] dy\right)$$

The proof is provided in section 8. Note that if $I = 0$, the exponential vanishes and this reduces to Theorem 3.

The number of unsuccessful mutations by time t/Nu_1s is of order $1/s$, so as in the discussion of the previous theorem, (A3) implies that we can ignore the possibility that the successful type 2 comes from a type 1 family that dies out. To explain the form of the limit, let $t' = t/(Nu_1s)$ to simplify notation. Successful type 1 mutations happen at rate Nu_1s. When one occurs at time r, it will cover a space-time volume at time t' of size

$$v(r) = \gamma_d c_d^d (t' - r)^{d+1}/(d+1)$$

and hence produce a Poisson number of successful type 2 mutations with mean $u_2 sv(r)$. A well-known and useful result, which can be found in Parzen (1999) or on page 421 of Komarova (2007), implies

$$P(\sigma_2 > t') = \exp\left(-\int_0^{t'} Nu_1 s \cdot [1 - \exp(-u_2 sv(r))]\, dr \right) \tag{12}$$ Parzen

This formula follows from thinning the Poisson process of successful type 1 mutations. If we only accept successful type 1 mutations that give rise to at least one successful type 2 mutation by time t', then the number of accepted points is Poisson with a mean equal to the integral in the formula. Since we have $\sigma_2 > t'$ if and only if no points are accepted, we have (12). The theorem follows easily from (12) by changing variables $y = Nu_1 s(t' - r)$ and using the definition of Γ in (8) to conclude

$$u_2 s\gamma_d c_d^d (Nu_1 s)^{-(d+1)} = \gamma_d/\Gamma$$

The careful reader will have noticed that we have assumed that the production of successful type 2's by different type 1 families are independent. We will prove this by showing that with high probability the space time cones that successful type 1 mutations generate are disjoint. Once this is done (12) will be proved and the Theorem follows.

3.3 $\Gamma \to \infty$

In this case there will be a large number of successful type 1 mutations before the first successful type 2 mutant occurs. A new feature in this case is that the first successful 2 may come from a type 1 family that dies out.

To state our result, we need several definitions. Let $\beta_2 = \pi$ and for $d \geq 3$ let β_d be the probability that two simple random walks started at 0 and $e_1 = (1, 0, \ldots 0)$ never hit. Let

$$\alpha_d = \begin{cases} 1 & d = 1 \\ 1/(2d\beta_d) & d \geq 2 \end{cases}.$$

When a type 2 mutation arises in a type 1 family that will die out, the type 2's have a fitness of s compared to the type 1's and of $2s$ compare to the type 0's, so its success probability is between s and $2s$. To avoid the difficulty of calculating the probability of success of a type 2 mutation landing at a random location on the space-time set occupied by an unsuccessful type 1 family, we will introduce a "fudge factor" $1 \leq \rho_2 \leq 2$ and assume that this probability is asymptotically $\rho_2 s$. Although the factor ρ_2 is annoying from a mathematical point of view, in practical applications our uncertainty about the mutation rates u_i is larger than uncertainty about the value of ρ_2.

Define $J = 1/u_2 \ell(s)$ and

$$K = \Gamma^{1/(d+2)} = (Nu_1 s)^{(d+1)/(d+2)} \cdot (c_d^d u_2 s)^{-1/(d+2)}. \tag{13}$$ Kdef

As the reader will see K is (the order of magnitude of) the number of successful type 1 mutations needed to get a successful type 2 mutation from a successful type 1 family. J is (the order of magnitude of) the number of successful type 1 mutations that will occur before we get a successful type 2 mutation from an unsuccessful type 1 family. Note that in the second statement we are measuring time in terms of the number of successful type 1 mutations, which occur roughly every $1/Nu_1s$ units of time. This somewhat awkward definition of J is needed for the following to be accurate.

lastth **Theorem 5.** *If $u_1 \le u_2$, $\Gamma \to \infty$, and (A3) holds then*

$$P(\sigma_2 > t/Nu_1s) \approx \exp\left(-\frac{\gamma_d (t/K)^{d+2}}{(d+1)(d+2)} - \rho_2\alpha_d(t/J)\right)$$

This informal statement is shorthand for three precise results. If $J/K \to \infty$ then unsuccessful type 1's can be ignored and

$$P(\sigma_2 > Kt/Nu_1s) \to \exp\left(-\frac{\gamma_d t^{d+2}}{(d+1)(d+2)}\right)$$

If $J/K \to 0$ then the successful type 2 is always born in an unsuccessful type 1 family and

$$P(\sigma_2 > Jt/Nu_1s) \to \exp\left(-\rho_2\alpha_d t\right)$$

If $K/(J+K) \to \theta \in (0,1)$ then both successful and unsuccessful type 1's contribute to the limit

$$P(\sigma_2 > (K+J)t/Nu_1s) \to \exp\left(-\frac{\gamma_d (t/\theta)^{d+2}}{(d+1)(d+2)} - \rho_2\alpha_d(t/(1-\theta))\right)$$

To prepare for their proofs note that if $u_1 = u_2$ then

$$K = N^{(d+1)/(d+2)}\left(\frac{us}{c_d}\right)^{d/(d+2)} = K_{pos}^{(d+1)/(d+2)} \tag{14}$$ KvsKpos

by (7). Thus if $u_1 \le u_2$ and K is large (as it is when $\Gamma \to \infty$) then $K \ll K_{pos}$ so σ_2 occurs before the 1's reach positive density.

4 Numerical Examples

sec:numex To help understand the conditions that specify the various regimes of behavior in our theorems, we will consider some concrete examples. We consider the case $d = 2$, which is relevant to carcinomas, i.e., cancers of the epithelial tissue which form linings of the body. Most cancers are of epithelial origin, including breast, colon, lung, esophageal, renal, and many others.

4.1 Phase diagram for $s = 0.01$ in $d = 2$

In a study of glioblastoma and colorectal cancer, Bozic et al (2010) concluded that the average selective advantage of somatic mutations was surprisingly small, 0.004. Here we adopt a slightly larger value, 0.01. In this subsection we will set $N = 10^c$, $u_1 = 10^{-a}$, and $u_2 = 10^{-b}$, where a, b, and c are all positive constants, and determine the regions in which our theorems can be applied. We will take $b = a - 2$, since in many cancers (e.g. breast, colorectal) an early mutational hit damages DNA replication machinery within the cell and leads to elevated mutation rates.

To identify the order of magnitude of these constants we note that cells have a diameter of roughly $10^{-5}\,m$ so there are 10^6 in 1 cm^2, and 10^8 in $(10\,cm)^2$. The point mutation rate has been estimated, see Jones et al. (2008), to be 5×10^{-10} per nucleotide per cell division. To compute the u_i this number needs to be multiplied by the number of nulceotides that when mutated lead to cancer. In some cases there are a small number of nonsynonymous mutations that achieve the desired effect, while in other cases there may be hundreds of possible mutations that knock out the gene and there may be a number of genes which can be hit. Bozic et al (2008) state that in their applications the number of possible mutations is 34,000. Thus mutation rates can range from 10^{-9} to 10^{-5}, or can be larger after the mechanisms that govern DNA replication are damaged.

To begin the study of our special case $s = 0.01$ and $d = 2$, we note that

$$\begin{aligned} c_2 = (s/\ln(1/s))^{1/2} = 0.0466 \qquad & \log_{10}(c_2) = -1.322 \\ \ell(s) = \ln(1/s)/s = 460.5 \qquad & \log_{10}(\ell(s)) = 2.663 \end{aligned}$$

Here $\log_{10}$ is the base 10 logarithm, and just in this one instance we have written ln rather than log for the natural logarithm.

In order to get started we need to check the technical conditions (A0) $(1/u_1) \gg \ell(s)^{(d+2)/2}$. To identify the boundaries between the various regimes we will replace $\ll$ by $<$. If we do this we need

$$a > 5.362$$

Condition (A1) says $(c_d/u_2 s)^{d/(d+1)} \ll N$. If we again replace $\ll$ by $<$, then the condition is

$$\left(0.0466 \cdot 10^{b+2}\right)^{2/3} < 10^c,$$

which after taking logarithms and simplifying can be written as

$$b - 3c/2 < -0.67. \tag{15}$$ aneqb_A1

We skip over (A2), which is $\Gamma \to 0$, because we will consider Γ later. Condition $(A3)$, which is $u_2 \ll 1/\ell(s)$ is now simply

$$b > 2.66. \tag{16}$$ aneqb_A3

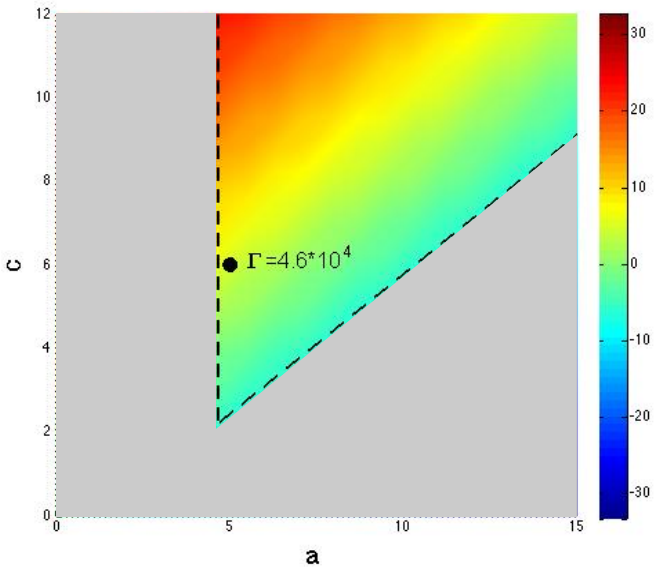

Figure 1: Parameter region delineated by assumptions $(A1)$ and $(A3)$, for $s = 0.01$, $d = 2$, and $b = a - 2$ (upper right region between the dashed lines). The vertical line is $(A3)$. The color at each point represents the value of $\log \Gamma$. The dot represents the parameters used in the example for colorectal cancer. newplot_ex

Figure 1 shows the parameter regime in which $(A1)$ and $(A3)$ apply (upper right region between the two dashed lines), when $s = 0.01$ and $d = 2$. To reduce to two dimensions we have assumed $b = a - 2$.

To determine the specific regimes in which Theorems 3, 4, and 5 hold, we must consider whether the parameter Γ tends to zero, infinity, or a positive real number. When $d = 2$ and $s = 0.01$

$$\begin{aligned} \Gamma &= (Nu_1 s)^{d+1}(c_d^d u_2 s)^{-1} \\ \log(\Gamma) &= 3(c - a - 2) - (-2.644 - b - 2) = 3(c - a) + b - 1.356 \end{aligned} \tag{17}$$ gammaeq

Figure 1 shows the value of $\log \Gamma$ at each point in the plane when $b = a - 2$. In contemplating the size of Γ it is important to remember that we are interested in $K = \Gamma^{1/(d+2)}$ which is approximately the number of successful mutations before σ_2 and its relationship to $J = 1/u_2\ell(s)$ which in our example has $\log J = b - 2.65$.

4.2 Application to colorectal cancer initiation

For a concrete example, we consider the process of cancer initiation in the sigmoid colon. The cells of the colon are subdivided into partially-isolated subpopulations of proliferative units, called *colonic crypts*. Each crypt is thought to contain approximately 4-20 stem cells which give rise to approximately 2000 differentiated cells

(Nicolas et al, 2007; Bach et al, 2000). These stem cells divide and may accumulate genetic mutations which spread until most stem cells within the crypt carry that mutation. Clonal expansion among crypts is possible via crypt bifurcation: a process during which a single crypt subdivides into two separate crypts via partitioning and regrowth of the stem and differentiated cell populations. If the stem cells within a crypt carry a mutation which is advantageous to cellular fitness, the rate of crypt bifurcation may be elevated, leading to clonal expansion of the number of crypts carrying the mutation.

Here, following the example of Martens et al (2011), we take each colonic crypt to be a single 'agent' in the model and consider a 'lattice' of crypts in the tissue of the sigmoid colon. The process of crypt bifurcation can then be thought of as analogous to the dynamics of the biased voter model. Specifically, when a crypt bifurcates we assume that it replaces a neighboring crypt. We consider the domain of the process to be the inner surface of the colon, which is a cylindrical structure; thus $d = 2$. Using the estimate of approximately 16 crypts per square mm (Cheng et al, 1984), we obtain an estimate of $N = 945000$ crypts in the entire sigmoid colon.

It has been observed that the incidence of colon cancer is higher in people who suffer from diseases of the colon such as ulcerative colitis and Crohn's disease. In such patients, the mutation rate u_1 per crypt cycle per crypt division time (in patients with predisposing conditions) is approximated to be 10^{-5} (thus $a = 5$), using baseline estimates of 10^{-10} for the mutation probability in a single nucleotide per cell division, ~ 10 target genes important in colorectal cancer initiation, and approximately 3 days between stem cell divisions in a crypt ((Totarfumo et al, 1987; see Martens et al, 2011 for a detailed description of this calculation). Here, the condition (A0) is not quite satisfied; however since we believe this condition can be improved, we will still proceed to see what these theorems can tell us.

It is widely accepted that a sequence of multiple mutations is necessary for the initiation of colon cancer. For example, a single defective allele of the gene *Adenomatous polyposis coli* (APC) can result in a condition in which the intestinal epithelium is studded with benign polyps. However, additional mutations (e.g. in genes p53 or kRAS) are required to initiate cancer. Furthermore, it has been suggested that inactivation of one APC allele may induce chromosomal aberrations (Ceol et al, 2007)); therefore we consider the case where $b = a - 2 = 3$.

In the following we are interested in understanding the dynamics of mutation accumulation in the sigmoid colon for patients with conditions such as ulcerative colitis and Crohn's disease. These conditions may often affect only a small portion of the colon; thus we consider a range of $N = 10^c$ for $c = 4, 5, 6$. Using Figure 1 we observe that for $a = 5, b = 3$, as the size of the affected part of the colon increases, the system dynamics change from the regime of Theorem 3 to the regime of Theorem 4 and Theorem 5. This provides us with some insight into how the spatial process of initiation may vary between individuals with differing severity of predisposing conditions, since predisposing conditions such as ulcerative colitis and

Crohn's disease cause inflammation and higher baseline mutation rates, as well as differing sizes of affected tissue regions.

We next consider the time for a crypt to acquire two hits once ulcerative colitis has already begun to affect crypt dynamics, and compare the model predictions to available epidemiological data. Although the extent of the affected area of the colon varies between patients at the time of diagnosis, a majority of patients are diagnosed with *distal* ulcerative colitis which is confined to the left side (approximately a third) of the colon. Thus we estimate of the number of crypts in the affected area as $N \approx 945000/3$. Recall that $\rho_2 s$ is the probability of success of a type-2 mutation landing at a random location on the space-time set occupied by an unsuccessful type-1 family; although its exact value is unknown, $\rho_2 \in [1, 2]$. For these parameters Theorem 5 holds, and we have $J = 2.17, K = 6.1$. Therefore the expression in Theorem 5 can be simplified by ignoring the term within the exponential involving K^{-4}. Using Theorem 5, we next calculate the probability of acquiring a successful type-2 cell within 10 years of developing ulcerative colitis. In colons affected by ulcerative colitis, it has been estimated that a crypt cycle takes approximately 100 days (Cheng et. al 1986). Since 10 years corresponds to approximately 36.5 crypt cycles, we find that $P(\sigma_2 < 36.5) \approx 0.041$ to 0.081, where the variation comes from the uncerstainty in ρ_2. It has been suggested that colorectal cancer develops as a result of two such rate-limiting genetic hits, followed by one additional high-frequency event (Luebeck and Moogavkar 2002). Thus we may compare this estimate to epidemiological observations of the progression rate to colorectal cancer within 10 years after developing ulcerative colitis, which is $\approx 2\%$ (Eaden et. al. 2001). Taking into account variability in extent of affected area at diagnosis as well as potential delays in the high frequency event and diagnosis time, we find the model prediction to be roughly consistent with available epidemiological observations. The model can then be utilized to provide further insights into risk stratification of the patient population using various factors such as extent of the affected area in colitis diagnoses, lifestyle and diet choices which may elevate mutation rates.

In followup works, we will utilize this model to study spatial measures of genetic diversity in premalignant tissue, as well as the phenomenon of field cancerization. Genetic diversity is an important marker of prognosis in premalignant conditions as well as tumor progression (Merlo et. al. 2006). For example, heterogeneity measures in a premalignant condition called Barretts esophagus were shown to strongly correlate with the likelihood of progression to esophageal cancer (Maley et. al., 2006). However, it is often difficult to accurately measure the amount of genetic heterogeneity in sequencing studies, and there is a need for improved spatial sampling guidelines. Field cancerization refers to the clinical observation that there is an increased likelihood for local recurrences in the same region of tissue after a primary tumor is surgically removed. This increased likelihood is suspected to be due to the existence of premalignant lesions at the time of diagnosis; however, in practice these lesions are usually undetectable and appear histologically normal. Thus, mathematical predictions of

spatial heterogeneity based upon fundamental tissue and cancer parameters will inform spatial sampling guidelines. In addition, these models can provide prognostic insights about the extent of premalignant lesions, probability of local recurrences, and the degree to which recurrent tumors are genetically related to primary tumors. These extensions will require the consideration of general k-step mutational pathways, where k varies depending on cancer type.

5 Construction and duality

sec:grep

To make the paper self-contained and to recall some facts that may not be widely known, we will now construct the two type biased voter model and explain its duality with coalescing branching random walk.

To construct the biased voter model, we follow the approach in Griffeath (1978). Associated with each order pair (x, y) of nearest neighbors, we have two Poisson processes, $T_n^{x,y,v}$, $n \geq 1$ and $T_n^{x,y,b}$, $n \geq 1$ with rates $1/2d$ and $(\lambda - 1)/2d$. Here, all of the Poisson processes are independent, and together constitute the *graphical representation*. At each time $t = T^{x,y,v}$ we draw an arrow $(y, t) \to (x, t)$ and put a δ at (x, t), while at each time $t = T^{x,y,b}$ we draw an arrow $(y, t) \to (x, t)$. We think of arrows as little tubes that allow fluid to flow in the direction indicated, while the δ's are dams that stop the passage of the fluid. The δ's occur just before the arrows so they don't block the fluid that flows through them.

Given an initial set A of sites that are occupied by 1's, the set of sites that are occupied by 1's at time t is the set ξ_t^A of points that are wet if fluid is injected at points of A at time 0. By checking cases, one can see that the effect of an arrow-δ from y to x is as follows:

before		after	
x	y	x	y
0	0	0	0
0	1	1	1
1	0	0	0
1	1	1	1

The first case should be clear. In the second the arrow spreads the fluid from y to x. In the third the δ at x stops the fluid, but there is nothing from y to replace it, while in the fourth case there is. Thus the arrow-δ produces a voter model step: x imitates y.

If we remove the δ, then only the third line changes and the overall result is a birth from y to x, with the two particles coalescing to one if x is already occupied:

before		after	
x	y	x	y
0	0	0	0
0	1	1	1
1	0	1	0
1	1	1	1

The graphical representation has the useful property that it constructs the biased voter model for all initial conditions on the same probability space. As Harris (1976) noted this implies that the constructed processes are additive:

$$\xi_t^{A\cup B} = \xi_t^A \cup \xi_t^B. \tag{18}$$ additive

since a space-time point can be reached from $A \cup B$ at time 0 if and only if it can be reached from A or from B. A consequence of additivity is that $A \to \xi_t^A$ is increasing, a property that is called "attractive."

An important reason for constructing a process from a graphical representation is that it allows us to construct a dual process. Let $\zeta_r^{x,t}$ be the set of points at time $t-r$ that can be reached by a path starting from x at time t that goes down the graphical representation and crosses the arrows in the direction OPPOSITE their orientation. If we recall that the δ's occur just before the arrows on the way up then we see that the effect of an arrow-δ from y to x on the dual process is

before		after	
x	y	x	y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	0	1

In words this is a coalescing random walk. If there is a particle at x (corresponding to a 1) it jumps to y. If there is also a particle at y the two coalesce to 1. It is easy to see that an arrow without a δ has the same effect in the dual as it did in the forward process except that now the birth is from x to y.

Given a set of sites B, let $\zeta_s^{B,t} = \cup_{x\in B}\zeta_s^{x,s}$. It is immediate from the definitions that

$$\{\xi_t^A \cap B \neq \emptyset\} = \{\zeta_t^{B,t} \cap A \neq \emptyset\}$$

i.e., the two events are equal. To get rid of the superscript t from the dual process, we note that if $t < t'$ then the distribution of $\zeta_r^{B,t'}$ for $r \le t$ is the same as $\zeta_r^{B,t}$ for $r \le t$. Invoking Kolmogorov's extension theorem there is process ζ_r^B defined for all time $r \ge 0$ that has the same distribution as $\zeta_r^{B,t}$ for $r \le t$. This process satisfies

$$P(\xi_t^A \cap B \neq \emptyset) = P(\zeta_t^B \cap A \neq \emptyset)$$

In what follows, we will be interested in the biased voter model with mutation $0 \to 1$ at rate u_1. Mutation can be incorporated into the graphical representation by adding independent Poisson processes $T_n^{x,\mu}$, $n \geq 1$ with rate u_1. If we let $\hat{\xi}_t^A$ be the biased voter model with mutation starting with A occupied at time t, and suppose that there are mutations at x_i at times $t_1 < t_2 < \ldots t_k < t$ then

$$\hat{\xi}_t^A = \xi_t^A \cup \xi_t^{x_1,t_1} \ldots \cup \xi_t^{x_k,t_k}$$

where $\xi_t^{x_i,t_i}$ is the biased voter model without mutation starting with x_i occupied at time t_i.

In our proofs, it will be useful to be able to quantify the notion that two processes, ξ_t^A and ξ_t^B or $\xi_t^{x_1,t_1}$ and $\xi_t^{x_2,t_2}$ are independent when they don't hit each other. To do this we use a coupling due to David Griffeath (1978). We define the first process on a graphical representation, and the second on an independent graphical representation with the caveat that events in the second process that involve an edge (x, y) where x or y is occupied in the first process must use the first graphical representation, so that the pair of processes has the same joint distribution as if they were both defined on the same graphical representation.

6 Results for the biased voter model

sec:bvm

6.1 Upper bounds

Our goal is to bound the size of type 1 families that later die out. The first step is to determine the effect on the process of conditioning it to die out. In the proofs in this section we will sometimes suppose in addition to $\lambda > 1$ that $\lambda \leq 2$ in order to get rid of λ's from the formulas.

ctodie **Lemma 6.1.** *Let ξ_t^0 is the set of 1's in a supercritical biased voter model with $\lambda = 1+s$ on $\mathbb{Z}^d$ starting from $\xi_0^0 = \{0\}$. Let T_0 be the time at which the process dies out. Let $\bar{\xi}_t^0$ be the biased voter model with $\bar{\xi}_0^0 = \{0\}$ and the roles of 1 and 0 interchanged, i.e., 1's give birth at rate 1, and 0's give birth at rate λ. Then*

$$(\{|\xi_t^0|, t \leq T_0\} | T_0 < \infty) =_d \{|\bar{\xi}_t^0|, t \leq T_0\}$$

Proof. If $\xi_t^0 = A$ with $|A| = k$ and $|\partial A| = \ell$ then $|\xi_t^0|$ grows to size $k+1$ at rate $\lambda\ell$, and shrinks to size $k-1$ at rate ℓ, so the transition probability of the embedded discrete time chain is

$$p(k, k+1) = \frac{\lambda}{1+\lambda} \qquad p(k, k-1) = \frac{1}{1+\lambda} \tag{19}$$ BVtp

If we let $\varphi(x) = \lambda^{-x}$ then it is easy to check that

$$\varphi(k) = p(k, k+1)\varphi(k+1) + p(k, k-1)\varphi(k-1)$$

hence if $a < x < b$ then

$$P_x(T_a < T_b) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)} \qquad P_x(T_b < T_a) = \frac{\varphi(x) - \varphi(a)}{\varphi(b) - \varphi(a)} \tag{20}$$ asymm

Let $a = 0$, $x = 1$, and $b \to \infty$ in the first formula

$$P_k(T_0 < \infty) = \lambda^{-k}. \tag{21}$$ hit0

If we condition a random walk with positive drift to hit 0 then the conditioned process has transition probability

$$\bar{p}(k, k+1) = \frac{p(k, k+1)\varphi(k+1)}{\varphi(k)} = \frac{1}{1+\lambda} \qquad \bar{p}(k, k-1) = \frac{\lambda}{1+\lambda} \tag{22}$$ CondTP

In words, the result is a random walk with the probabilities of up and down interchanged. Conditioning ξ_t to hit 0 does not change the exponential holding times, and desired result follows. □

The next three results are numerical upper bounds. As in part I these are based on the fact that when k is large, the size of the boundary in the biased voter model

$$\partial(k) \sim \begin{cases} 2d\beta_d k & d \geq 3 \\ 4\beta_2 k/\log k & d = 2. \end{cases} \tag{23}$$ Bdrysize

When $s = 0$ this follows from (I1) on page 202 of Cox, Durrett, and Perkins (2002). Also see the discussion of this point in Section 2 of Part I.

To extend this to the subcritical biased voter model we will use Girsanov's formula:

LetP_0 be the law of the voter model in which a 1 next to a 0 forces it to flip at rate $(1 + s/2)/2d$ and vice versa a 0 next to a 1 forces it to flip at rate $(1 + s/2)/2d$.

Let P_s be the law of a subcritical voter model in which a 1 next to a 0 forces it to flip at rate $1/2d$ and vice versa a 0 next to a 1 forces it to flip at rate $(1 + s)/2d$.

The speed up in P_0 is a harmless linear transformtion of the time scale, but it makes the rates at which things happen $= (2+s)/2d$ on each discordant edge. Thus when we look at the Radon-Nikodym derivative it will only depend on the embedded chain.

$$\frac{dP_s}{dP_0} = \left(\frac{2}{2+s}\right)^{n_+} \left(\frac{2+2s}{2+s}\right)^{n_-} \tag{24}$$ RN1

where n_+ is the number of up jumps and n_- is the number of down jumps. Rewriting the RN derivative we have

$$\frac{dP_s}{dP_0} = \left(\frac{2}{2+s}\right)^{n_+ - n_-} \left(\frac{4+4s}{4+4s+s^2}\right)^{n_-} \leq 1 \tag{25}$$ RN2

so events that have small probability under P_0 (e.g., the boundary size deviating from the stated formula) also have small probability under P_s.

Let $q(k) = (1 + \lambda)\partial(k)$ be the rate at which jumps occur when the subcritical biased voter model has k points.

ubdead **Lemma 6.2.** *Let $\bar{\xi}_t$ be the subcritical biased voter model defined in Lemma 6.1.*

$$g(k) \equiv E_k\left(\int_0^{T_0} |\bar{\xi}_t|\, dt\right) \le Ck\ell(s)$$

where $\ell(s)$ was defined in (2)

Proof. Due to the additivity property of the biased voter model (18) it suffices to prove the result when $k = 1$. Let $\bar{S}_n$ be the simple random walk that jumps according to $\bar{p}$ defined in (22) and note that $|\bar{\xi}_t|$ is a continuous time version of $\bar{S}_n$ that jumps approximately at rate $q(k)$ when in state k and makes jumps according to $\bar{p}$. Let $T_k^+ = \min\{n \ge 1 : S_n = k\}$. By considering the expected number of visits to k and the amount of time spent there on each one we have

$$E_1\left(\int_0^{T_0} |\bar{\xi}_t|\, dt\right) = \sum_{k=1}^{\infty} \frac{\bar{P}_1(T_k < T_0)}{\bar{P}_k(T_k^+ = \infty)} \cdot \frac{k}{q(k)} \tag{26}$$ deadmh

If $\bar{P}$ is the law of the conditioned chain with transition probability (22), and P is the law of the process with transition probability chain following (19) then by symmetry and (20),

$$\bar{P}_1(T_k < T_0) = P_{k-1}(T_0 < T_k) = \frac{\lambda^{-k} - \lambda^{-(k-1)}}{\lambda^{-k} - 1} = \frac{\lambda - 1}{\lambda^k - 1}. \tag{27}$$ P1Tk

Using symmetry and (20) again,

$$\begin{aligned} \bar{P}_k(T_k^+ = \infty) &= \frac{\lambda}{1+\lambda}\bar{P}_{k-1}(T_0 < T_k) = \frac{\lambda}{1+\lambda}P_1(T_k < T_0) \\ &= \frac{\lambda}{1+\lambda} \cdot \frac{\lambda^{-1} - 1}{\lambda^{-k} - 1} = \frac{\lambda - 1}{(1+\lambda)(1-\lambda^{-k})}, \end{aligned} \tag{28}$$ PkTk+

so we have

$$\frac{\bar{P}_1(T_k < T_0)}{\bar{P}_k(T_k^+ = \infty)} = \left(\frac{\lambda - 1}{\lambda^k - 1}\right)\left(\frac{1+\lambda}{\lambda - 1}\right)\lambda^{-k}(\lambda^k - 1) = \lambda^{-k}(1+\lambda)$$

and (26) becomes

$$E\left(\int_0^{T_0} |\bar{\xi}_t|\, dt\right) = (1+\lambda)\sum_{k=1}^{\infty} \lambda^{-k} \cdot \frac{k}{q(k)}. \tag{29}$$ deadmh2

In one dimension $q(k) = 2(1+\lambda)$, so doing some algebra and using the formula for the mean of the geometric distribution, two times the quantity in (29) is

$$\begin{aligned} \sum_{k=1}^{\infty} k\lambda^{-k} &= \frac{1/\lambda}{(1-1/\lambda)} \sum_{k=1}^{\infty} k\lambda^{-(k-1)}(1 - 1/\lambda) \\ &= \frac{1/\lambda}{(1-1/\lambda)^2} \le Cs^{-2}. \end{aligned} \tag{30}$$ a3bd1

Since $q(k) \sim 2d(1+\lambda)\beta_d k$, in $d \geq 3$, so the quantity in (29) is

$$\frac{1}{2d\beta_d} \sum_{k=1}^{\infty} \lambda^{-k} = \frac{1}{2d\beta_d} \cdot \frac{1/\lambda}{1-1/\lambda} \leq Cs^{-1}. \tag{31}$$ a3bd3

In $d = 2$ we have $q(k) \sim 4(1+\lambda)\beta_2 k/\log k$ so the quantity in (29) is

$$\frac{1}{4\beta_2} \sum_{k=1}^{\infty} \lambda^{-k} \log k \leq Cs^{-1} \log(1/s). \tag{32}$$ a3bd2

To see the last inequality note that for $k \leq 1/s^2$, $\log k \leq 2\log(1/s)$ and as $s \to 0$ we can ignore the contribution from $k > 1/s^2$. □

In order to conclude that the limit of the expected values is the expected value of the limit we need a bound for the second moment. We begin with the case $k = 1$.

mom2d **Lemma 6.3.** *Let $\bar{\xi}_t^0$ be the subcritical biased voter model defined in Lemma 6.1.*

$$E_1 \left(\int_0^{T_0} |\bar{\xi}_t^0| \, dt \right)^2 \leq C\ell(s)^2/s.$$

Proof. Using the Markov property

$$\begin{aligned} E_k \left(\int_0^{T_0} |\bar{\xi}_t^0| \, dt \right)^2 &= 2E_k \int_0^{T_0} dr \, |\bar{\xi}_r^0| \int_r^{T_0} |\bar{\xi}_t^0| \, dt \qquad (33) \quad \text{m2occ} \\ &= 2E_k \int_0^{T_0} dr \, |\bar{\xi}_r^0| g(|\bar{\xi}_r^0|) \leq 2C\ell(s) E_k \int_0^{T_0} dr \, |\bar{\xi}_r^0|^2 \end{aligned}$$

by Lemma 6.2. Thus by the reasoning that lead to (29)

$$E_1 \left(\int_0^{T_0} |\bar{\xi}_t^0| \, dt \right)^2 \leq 2C\ell(s) \cdot (1+\lambda) \sum_{k=1}^{\infty} \lambda^{-k} \cdot \frac{k^2}{q(k)}. \tag{34}$$ deadmh3

The remainder of the proof is similar to the previous argument

$$\begin{aligned} d = 1 \qquad & q(k) = 2(1+\lambda) \qquad && \frac{1}{2} \sum_{k=1}^{\infty} k^2 \lambda^{-k} \leq Cs^{-3} \\ d \geq 3 \qquad & q(k) \sim 2d(1+\lambda)\beta_d k \qquad && \frac{1}{2d\beta_d} \sum_{k=1}^{\infty} k\lambda^{-k} \leq Cs^{-2} \\ d = 2 \qquad & q(k) \sim 4(1+\lambda)\beta_2 \frac{k}{\log k} \qquad && \frac{1}{4\beta_2} \sum_{k=1}^{\infty} \lambda^{-k} k \log k \leq Cs^{-2} \log(1/s). \end{aligned}$$

Recalling the definition of $\ell(s)$ in (2), we see that the right-hand side is always $C\ell(s)/s$ and we have proved the desired result. □

ub2dead **Lemma 6.4.** *Let $\bar{\xi}_t$ be the subcritical biased voter model defined in Lemma 6.1.*

$$E_j\left(\int_0^{T_0} |\bar{\xi}_t|\,dt\right)^2 \le \frac{C}{s}\cdot\frac{j^3}{q(j)} + C(\lambda^j-1)\left(\frac{\ell(s)}{s}\right)^2.$$

Note that when $j = a/s$ and $a \le 1$ both terms on the right are bounded by $Ca(\ell(s)/s)^2$.

Proof. Combining (33) with (26) we have

$$E_j\left(\int_0^{T_0} |\bar{\xi}_t|\,dt\right)^2 \le 2C\ell(s)\sum_{k=1}^{\infty}\frac{\bar{P}_j(T_k<T_0)}{\bar{P}_k(T_k^+=\infty)}\cdot\frac{k}{q(k)}. \tag{35}$$ deadm2

When $j \ge k$, $\bar{P}_j(T_k < T_0) = 1$ while for $j \le k$, (27) becomes

$$\bar{P}_j(T_k<T_0) = P_{k-j}(T_0<T_k) = \frac{\lambda^{-k}-\lambda^{-(k-j)}}{\lambda^{-k}-1} = \frac{\lambda^j-1}{\lambda^k-1}. \tag{36}$$ PjTk

The formula in (28) has not changed

$$\bar{P}_k(T_k^+=\infty) = \frac{\lambda-1}{(1+\lambda)(1-\lambda^{-k})}$$

so we have

$$h(k) \equiv \frac{\bar{P}_j(T_k<T_0)}{\bar{P}_k(T_k^+=\infty)} = \begin{cases} (1+\lambda)\dfrac{1-\lambda^{-k}}{\lambda-1} & k \le j \\ (1+\lambda)(\lambda^j-1)\dfrac{\lambda^{-k}}{\lambda-1} & k \ge j. \end{cases}$$

To bound the first part of the sum, $j \le k$ in (35), we begin by noting that $\lambda - 1 = s$ and $(1+\lambda)(1-\lambda^{-k}) \le 3$ so $h(k) \le 3/s$. At this point there are three cases $q(j) = O(1)$, $q(j) = O(j/\log(j))$, and $q(j) = O(j)$ for $d = 1, 2$ and $d \ge 3$. However in each case $k^2/q(k)$ is increasing so

$$\sum_{k=1}^{j} h(k)\frac{k^2}{q(k)} \le \frac{C}{s}\cdot j\cdot\frac{j^2}{q(j)}.$$

This is the first term on the right-hand side of the lemma. For the second part of the sum, we use

$$\sum_{k=j}^{\infty} h(k)\frac{k^2}{q(k)} \le C\frac{\lambda^j}{s}\sum_{k=j}^{\infty}\lambda^{-k}\frac{k^2}{q(k)}$$

so using the computation for (34) in the previous lemma gives the second term. □

6.2 Limit theorems

Our first step is to generalize Lemma 1 from part I. Let T_k be the first time $|\xi_t^0| = k$ and let

$$a(n) = \begin{cases} n^2 & d = 1, \\ n \log n & d = 2, \\ n & d \geq 3. \end{cases}$$

(This time there is no 2 in the definition for $d = 2$.)

dead1lb **Lemma 6.5.** *Let $\epsilon > 0$ and let ξ_t^n be the biased voter model with $\lambda = 1-1/n$. Writing $[x]$ for the integer part we have*

$$\left(\frac{|\xi^n_{T_{[n\epsilon]}+a(n)t}|}{n} \middle| T_{[n\epsilon]} < \infty \right) \Rightarrow (Y_t | Y_0 = \epsilon), \tag{37}$$ procconv

where $\Rightarrow$ indicates convergence in distribution of the stochastic processes as $n \to \infty$.

$$dY_t = \begin{cases} -2\,dt + 2\,dB_t & d = 1, \\ -2d\beta_d Y_t\,dt + \sqrt{4d\beta_d Y_t}\,dW_t & d \geq 2. \end{cases}$$

In $d = 1$ the process is stopped when it hits 0. In $d \geq 2$, 0 is an absorbing boundary so we don't have to stop the process.

Proof. In $d = 1$ the result is trivial. The size of the set increases by 1 at rate $2(1-1/n)$ and decreases by 1 at rate 2, so if $\xi^n(n^2 t) = k$

$$\begin{aligned} \text{infinitesimal mean} &= 2n^2 \cdot \frac{1}{n} \cdot \frac{-1}{n} = -2 \\ \text{infinitesimal variance} &= 2n^2 \cdot \frac{1}{n^2} \cdot \left(2 - \frac{1}{n}\right) \to 4. \end{aligned}$$

In this and the next two calculations the first factor is the time scaling, the second comes from the fact that jumps change the scaled process by $1/n$. The third term is the difference of the rates in the first case and the sum in the second.

The one-dimensional case is easy because if we start from a single occupied site the size of the boundary, i.e., number of $0-1$ edges is always 2 until the process dies out. Using the formulas for the boundary size in (23), we see that in $d \geq 3$ if $\xi^n(nt) = k$ with $k/n = x$ then

$$\begin{aligned} \text{infinitesimal mean} &= n \cdot \frac{1}{n} \cdot \frac{-2d\beta_d k}{n} \to -2d\beta_d x \\ \text{infinitesimal variance} &= n \cdot \frac{1}{n^2} \cdot \frac{2d\beta_d k}{n} \cdot \left(2 - \frac{1}{n}\right) \to 4d\beta_d x \end{aligned}$$

while in $d = 2$ if $\xi^n(tn \log n) = k$ with $k/n = x$ then

$$\begin{aligned} \text{infinitesimal mean} &= n \log n \cdot \frac{1}{n} \cdot \frac{-4\beta_d k}{n \log k} \cdot \to -4\beta_d x \\ \text{infinitesimal variance} &= n \log n \cdot \frac{1}{n^2} \cdot \frac{4\beta_d k}{\log k} \cdot \left(2 - \frac{1}{n}\right) \to 8\beta_d x. \end{aligned}$$

Having shown the convergence of infinitesimal mean and variance to that of a stochastic differential equation with a well-posed martingale problem, the result follows. See e.g., Theorem 4.1 on page 354 in Ethier and Kurtz (1986). □

In $d \geq 2$, Lemma 6.5 can be extended to a measure valued limit. We begin by describing the limit. A measure valued process X_t is a super-Brownian motion with branching rate γ, diffusion coefficient σ^2 and drift θ if it is a solution of the following martingale problem. Let Δ denote the Laplacian, and use $\mu(f)$ to denote the integral of the function f with respect to the measure μ. For all $\phi \in C_K^\infty(\mathbb{R}^d)$ (smooth functions with compact support)

$$Z_t(\phi) = X_t(\phi) - X_0(\phi) - \int_0^t X_s(\sigma^2 \Delta\phi/2 + \theta\phi)\, ds$$

is a martingale with variance process

$$\langle Z(\phi) \rangle_t = \int_0^t X_s(\gamma\phi^2)\, ds.$$

mvlim **Lemma 6.6.** *Suppose $d \geq 2$ and let X_t^n be the measure that assigns mass $1/n$ to each point $a(n)^{-1/2}\xi^0_{T_{n\epsilon}+a(n)t}$. If there exists a subsequence $n(k)$ such that $X_0^{n(k)}$ converges weakly to a limit, then $(X^{n(k)}|T_{n\epsilon} < \infty)$ converges weakly to a super-Brownian motion with branching rate $2\beta_d$, diffusion coefficient 1, and drift $-\beta_d$.*

We mention this result because we think it is interesting. Unfortunately it does not give us what we need to prove Lemma 6.7, so we will not use it in what follows.

Why is this true? For $d \geq 3$ this is a special case of Theorem 1.3 in Cox and Perkins (2005) since the biased voter model is a special case of the voter model perturbation they consider. The situation is not as clean in the more difficult case of $d = 2$. To quote Ed Perkins, it is an easier argument than the Lotka-Volterra models considered in Theorem 1.2 of Cox and Perkins (2008). □

ubst **Lemma 6.7.** *Let $\delta > 0$. If M is large then the probability an unsuccessful type 1 family will last for time $\geq M\ell(s)$ or will escape from a cube of radius $M\ell(s)^{1/2}$ is $\leq \delta s$.*

In $d = 1$ the biased voter model is an interval when it is not empty so the desired result follows from Lemma 6.5. To see this recall that in that lemma $s = 1/n$ and that $\ell(1/n) = a(n)$.

To prove this in $d \geq 2$ we will combine the first half of Theorem 4 from Bramson, Cox, and LeGall (2001) with the trivial fact that the voter model dominates the subcritical biased voter model. Their result concerns the ordinary voter model with kernel $p(x, y)$. That is, voter at x changes opinions at rate 1, and imitates the one at y with probability $p(x, y)$ where $p(x, y) = p(0, y - x)$ is irreducible and symmetric with $p(0, 0) = 0$ and $\sum_x p(0, x)x_i x_j = \sigma^2\delta(i, j)$. Here $\delta(i, j) = 1$ if $i = j$ and 0 otherwise. To get a limit, we scale space so that the voters live on $\mathbb{Z}^d/\sqrt{n}$, run time at rate n and denote the resulting voter model by ξ_t^n. Let $m_n = n/\log n$ in $d = 2$ and n in $d \geq 3$, and define a measure valued process by

$$X_t^n = \frac{1}{m_n} \sum_{y \in \xi_t^n} \delta_y.$$

We write $X_t^{n,0}$ when the initial state is $\xi_0^n = \{0\}$. Let $\mathbf{D}$ be the space of functions from $[0, \infty)$ into the space of finite measures on $\mathbb{R}^d$ that are continuous in the weak topology.

BCL **Theorem 6.** *Assume $d \geq 2$ and let $\mathbf{N}_0$ be the excursion measure of super-Brownian motion on $\mathbb{R}^d$ with branching rate $2\beta_d$ and diffusion coefficient σ^2.*
Let $\alpha > 0$ and let F be a bounded continuous function on $\mathbf{D}$ with $F(\omega) = 0$ if $\omega_t = 0$ for $t \geq \alpha$. Then

$$\lim_{n \to \infty} m_n E F(X_\cdot^{n,0}) = \mathbf{N}_0(F).$$

The excursion measure is defined by starting the super process from $\epsilon\delta_0$, multiplying the probability measure by $1/\epsilon$ and letting $\epsilon \to 0$. See Section 3 of Bramson, Cox, and LeGall (2001) and references therein for more details. This is the super-process analogue of starting Brownian motion at ϵ, killing it when it hits 0, considering the limit of $(1/\epsilon)$ times the probability measure, which defines Ito's excursion measure. See Chapter XII of Revuz and Yor (1991) for a thorough treatment. In most cases the killed Brownian motion $\bar{B}_t$ dies out quickly but when $\epsilon < 1$

$$(1/\epsilon) P_\epsilon(\max_t \bar{B}_t > 1) = 1.$$

Theorem 6 was used in Merle (2008) to study the likelihood of the voter model wandering far away from its starting point. We will use a result of his to establish Lemma 6.7.

Proof of Lemma 6.7. We will first establish the result for the voter model. In particular, define T_0 to be the extinction time of the voter model started from a single site

on $\mathbb{Z}^d$, and $p(t) = P(T_0 > t)$. Recall from Bramson and Griffeath (1980a) the large t asymptotics

$$p(t) \sim \begin{cases} \frac{1}{\beta_d t}, & d \geq 3 \\ \frac{1}{\pi}\frac{\log t}{t}, & d = 2. \end{cases} \tag{38}$$

eq:ext_time

Based on these asymptotics it is easy to verify that

$$p\left(M\ell(s)\right) \sim s/M,$$

which establishes the claim with regards to survival time. In order to establish the claim with regard to escape from the large box we use Claim 1 of Merle 2008 (p828):

Claim 1. *There exists a positive K_0 and K_2 such that for any $\alpha > 1$ and any $A \geq 1$*

$$P\left(\sup_{t\leq 2\alpha}\sup_{x\in\xi_t^0}|x| > A\sqrt{\alpha}\right) \leq K_0 p(\alpha)\exp(-K_2 A)$$

where ξ_t^0 is a voter model started with a single seed at the origin.

Now consider

$$\begin{aligned} P\left(\sup_{t>0}\sup_{x\in\xi_t^0}|x| > M\sqrt{\ell(s)}\right) &\leq P\left(\sup_{t\leq 2\sqrt{M}\ell(s)}\sup_{x\in\xi_t^0}|x| > M\sqrt{\ell(s)}\right) + p\left(2\sqrt{M}\ell(s)\right) \\ &\leq p\left(2\sqrt{M}\ell(s)\right)\left(K_0 e^{-K_2\sqrt{M}} + 1\right). \end{aligned}$$

The result then follows for the subcritical biased voter model by comparison. □

7 Proof of Theorem 2

In this section we will show that if σ_1 is the birth time of the first successful type 1 then $P(\sigma_1 > t/Nu_1 s) \to e^{-t}$. To do this, we consider the space-time set G_r of $(x,q) \in \mathbb{Z}^d \times [0,r]$ with $||x||_\infty \leq L$ so that the biased voter model (constructed on the graphical representation described in Section 5) started from x at time q survives in the sense that it reaches size C_s/s where $C_s \to \infty$ as $s \to 0$. As explained in Section 2.2, the success probability is $\sim s$, as $s \to 0$.

Let $R = t/Nu_1 s$ and recall $N = L^d$. Let $|G_R| = \int_0^R |\{x : (x,q) \in G_R\}|\, dq$. Our goal is to show that

$$|G_R|/NRs \to 1 \quad \text{in probability.}$$

Since mutations to type 1 occur at rate u_1 and one that lands on G_R produces a success, the desired result will follow.

To approximate G_R, we will consider G_R^M the set of points (x,q) with $q \leq R$ so that the biased voter model escapes from the space-time set

$$(x,q) + ([-M\ell(s)^{1/2}, M\ell(s)^{1/2}]^d \times [0, M\ell(s)]),$$

where $\ell(s)$ was defined in (2). Let $\delta > 0$. It follows from Lemma 6.7 that if M is large enough then

$$P((x,q) \in G_R^M - G_R) \le \delta s. \tag{39}$$ GMvsG

Let $p_M = P((x,q) \in G_R^M)$ and note that $|p_M - s| < \delta s$.

To get bounds on the size of $|G_R^M|$, we note that if $\|x - x'\|_\infty > 2M\ell(s)^{1/2}$ or $|q - q'| > M\ell(s)$ then the events $(x,q) \in G_R^M$ and $(x',q') \in G_R^M$ are independent. From this it follows that

$$\begin{aligned}\text{var}\,(|G_R^M|) &= \sum_{x:||x||_\infty \le L} \sum_{x':||x'||_\infty \le L} \int_0^R dq \int_0^R dq' \,\text{cov}\,(1_{(x,q)\in G_r^M}, 1_{(x',q')\in G_r^M}) \\ &\le NR \cdot (4M\ell(s)^{1/2})^d \cdot 2M\ell(s) \cdot p_M = cM^{d+1} \cdot NRp_M \cdot \ell(s)^{(d+2)/2}.\end{aligned}$$

We need to show that $\text{var}\,(|G_R^M|) \ll (E|G_R^M|)^2$ where $(E|G_R^M|)^2 = (NRp_M)^2$. However, this follows from (A0) $(1/u_1) \gg \ell(s)^{(d+2)/2}$, and hence Chebyshev's inequality gives

$$|G_R^M|/NRp_M \to 1 \quad \text{in probability.}$$

Using (39) now we see that $P(|G_R^M - G_R| > k\delta sNR) \le 1/k$. Since δ is arbitrary, it follows that $|G_R|/NRs \to 1$ and the proof of Theorem 2 is complete.

Remark 1. *For all scenarios under consideration in this paper we have that at the time σ_2 the type 1 cells are at an asymptotically negligible density, this is enforced by either assumption (A1) or* (14)*. Note that the previous proof of Theorem 2 carries over identically if we assume that there are previous existing type 1 cells at asymptotically negligible density. Therefore, the sequence of scaled successful type 1 arrival times until σ_2 are asymptotically a Poisson process.*

8 Proofs of Theorems 3 and 4

sec:pfDM

Proof of Theorem 3. Let A_t be the event that the first successful type 2 mutant comes from an unsuccessful type 1 family that arises before time $t/Nu_1 s$. The expected number of such families is t/s. Using Lemma 6.2 to bound the space time volume covered by an unsuccessful mutation and ignoring the possibility that they overlap, we have

$$P(A_t) \le u_2 s \cdot \frac{t}{s} \cdot C\ell(s) \to 0 \tag{40}$$ notdead1

by (A3).

Under the simplified model, the total space-time volume occupied by descendants of the successful type 1 mutation up to time $\sigma_1 + t$ is

$$g(t) = \frac{\gamma_d c_d^d t^{d+1}}{d+1},$$

assuming that at this time the diameter is $< L$. Let $B(t)$ be the event that there is no successful type 2 mutation by time $\sigma_1 + t$, and let

$$t_2 = (u_2 s \gamma_d c_d^d/(d+1))^{-1/(d+1)}.$$

Using the Poisson approximation

$$P(B(Kt_2)) \sim \exp(-u_2 s g(Kt_2)) \to \exp(-K^{d+1}).$$

This shows that if K is large, $\sigma_2 - \sigma_1 \le Kt_2$ with high probability. Now the diameter of the ball covered by the descendants of the successful type 1 at time $\sigma_1 + Kt_2$ is

$$O(c_d t_2) = O((c_d/u_2 s)^{1/(d+1)}) \ll L$$

by assumption (A1), so our computation of the volume is legitimate. Finally, the time

$$t_2 = O((c_d^d u_2 s)^{-1/(d+1)}) \ll 1/N u_1 s$$

by assumption (A2). Since σ_1 is $O(1/Nu_1 s)$ this shows that the time difference $\sigma_2 - \sigma_1$ can be ignored. Since mutations to type 1 come at times of a Poisson process with rate $Nu_1 s$ this tells us that the possibility of a second successful type 1 mutation can also be ignored. □

Proof of Theorem 4. The only remaining detail is to show that with high probability the space-time cones generated by different type 1 mutations are dsjoint. This follows from the proof of Lemma 9.1. □

9 Proof of Theorem 5

sec:pfJK

There are three cases corresponding to the three limit theorems stated after the result. Define $Z_1^*(r)$ to be the number of descendants of successful type i mutations at time r and define $Z_1^0(r)$ to be the number of descendants of unsuccessful type 1 mutations at time r.

9.1 $J \gg K$

In this case the successful type 2 will come from a successful type 1 family. To simplify notation define

$$H_1(t) = \frac{\gamma_d}{u_2 s} \frac{t^{d+2}}{(d+1)(d+2)}, \quad \text{and} \quad T_1(t) = \frac{Kt}{N u_1 s}.$$

Since successful type 2 mutations occur at rate $u_2 s$ the desired result follows immediately from

s2area **Lemma 9.1.** *Under the assumptions of Theorem 5 if $J/K \to \infty$ then*

$$\frac{1}{H_1(t)} \int_0^{T_1(t)} Z_1^0(r)\, dr \to 0 \quad \textit{and} \quad \frac{1}{H_1(t)} \int_0^{T_1(t)} Z_1^*(r)\, dr \to 1$$

in probability as u_1, u_2 and s go to 0.

Proof. The first thing to show is that the contribution from unsuccessful type 1 mutations is negligible. The expected total number of unsuccessful type 1 mutations by time $T_1(t)$ is approximately tK/s. Therefore by Lemma 6.2 their expected total space-time contribution is $\le (tK/s) \cdot C\ell(s)$. Recalling that $J = 1/u_2\ell(s)$,

$$K \ll \frac{1}{u_2\ell(s)} \quad \text{implies} \quad \frac{K\ell(s)}{s} \ll \frac{1}{u_2 s}$$

and the successful type 2 mutation will not come from an unsuccessful type 1.

The next step is to show that we can assume that the volumes covered by distinct successful type 1 families are disjoint. By the light-cone argument used to derive (6), the volume A_T covered by type 1 families up to time T has

$$E(A_T) = N \int_0^T 1 - \exp(-\lambda_t)\, dt \quad \text{with} \quad \lambda_t = u_1 s \cdot \gamma_d c_d^d t^{d+1}/(d+1)$$

while if B_T is the volume covered by at least two successful type 1 families

$$E(B_T) = N \int_0^T 1 - \exp(-\lambda_t)(1+\lambda_t)\, dt.$$

The discussion before (14) implies that σ_2 occurs when 1's have low density, so $\lambda_{T_1(t)} \to 0$. Since $e^{-x} \ge 1 - x$, we always have $1 - e^{-x}(1+x) \le x^2$. In the other direction, if δ is small and $0 < x < \delta$, $1 - e^{-x} \ge x/2$. Combining our results we see that

$$E(B_{T_1(t)}) \le 2\lambda_{T_1(t)} E(A_{T_1(t)})$$

so Markov's inequality implies that overlaps can be ignored.

Recall that Theorem 2 tells us that successful type 1 mutations happen at times $0 < t_1 < t_2 < \dots$ of a Poisson process with rate $Nu_1 s$. Note that if we condition on the number M of mutations that have occurred by time $Kt/Nu_1 s$ then $\{t_1, t_2, \dots t_M\}$ has the same distribution as $\{v_1, v_2, \dots v_M\}$ where the v_i are independent random variables uniform on $[0, Kt/Nu_1 s]$. Let X_i be the space-time volume covered by the type 1 family starting at v_i, then

$$E[X_i | v_i] = \frac{\gamma_d c_d^d}{d+1} \left(\frac{Kt}{Nu_1 s} - v \right)^{d+1}.$$

Since v_i is uniform,

$$EX_i = \frac{\gamma_d c_d^d}{(d+1)(d+2)} \cdot \left(\frac{Kt}{Nu_1 s}\right)^{d+1} \tag{41}$$ EXi

where γ_d is the geometric constant defined in (5). The random sum $S_M = X_1 + \cdots + X_M$ has

$$ES_M = EX_i EM \tag{42}$$ meanrsum

$$\text{var}\,(S_M) = EM \text{var}\, X_i + \text{var}\,(M)(EX_i)^2 = EMEX_i^2 \tag{43}$$ varrsum

since M is Poisson and hence has $EM = \text{var}\, M$. To get an upper bound on EX_i^2, suppose the mutation occurs at time 0, and replace the cone by a cylinder to get

$$X_i \le \gamma_d \left(c_d \frac{Kt}{Nu_1 s}\right)^d \frac{Kt}{Nu_1 s}.$$

Since $EM = Kt$, using (41) and (43) it follows that

$$\text{var}\,(S_M) \le CKt \left(\frac{Kt}{Nu_1 s}\right)^{2d+2} = C\frac{(ES_M)^2}{Kt}.$$

Since $K \to \infty$, Chebyshev's inequality implies $S_M/ES_M \to 1$ in probability. Using (41) and (42)

$$ES_M = \frac{\gamma_d c_d^d}{(d+1)(d+2)} \cdot \frac{(Kt)^{d+2}}{(Nu_1 s)^{d+1}} = H_1(t) \tag{44}$$ ESM

where the last equality follows from the definition of K in (13).

The last result gives asymptotics for the volume covered by successful type 1 families, when we have ignored the possibility of further type 2 mutations. Now unsuccessful type 2 mutations occur at rate u_2 and have expected space time volume $\ell(s)$. Assumption (A3) implies $u_2\ell(s) \to 0$ so the loss of volume can be ignored. □

9.2 $J \ll K$

In this case the successful type 2 will come from an unsuccessful type 1 family. Let

$$H_2(t) = \frac{\alpha_d t}{u_2 s}, \quad \text{and} \quad T_2(t) = \frac{tJ}{Nu_1 s}.$$

As in the previous result it suffices to show

d2area **Lemma 9.2.** *Under the assumptions of Theorem 5 if $J/K \to 0$ then*

$$\frac{1}{H_2(t)} \int_0^{T_2(t)} Z_1^0(r)\, dr \to 1 \quad \textit{and} \quad \frac{1}{H_2(t)} \int_0^{T_2(t)} Z_1^*(r)\, dr \to 0,$$

in probability as u_1, u_2 and $s \to 0$.

The second part of the result follows easily from Lemma 9.1. To see this note that $H_i(t) = c_i(d,t)/u_2 s$, while the earlier calculation, see (44), implies

$$E\int_0^{T_2(t)} Z_1^*(r)\,dr \sim (J/K)^{d+2} H_1(t) = o(H_2(t)).$$

To prove the first part of this result we need some information about unsucccessful type 1 families. The probability that a subcritical voter model ξ_t^0 with $\lambda = 1-s$ hits ϵ/s is using (20)

$$\frac{(1-s)^{-1}-1}{(1-s)^{-\epsilon/s}} \approx \frac{s}{e^\epsilon - 1} \tag{45}$$ reach

Taking $\epsilon = 1$ in the last result, $n = 1/s$ in (37), and noting $a(1/s) = \ell(s)$ we see that type 1 families reach $1/s$ with probability s, and their total man-hours before extinction is approximated by

$$\frac{\ell(s)}{s}\left(\int_0^{T_0} Y_s\,ds \,\middle|\, Y_0 = 1\right).$$

Thus contributions to the mean that come from type 1 families that reach size $1/s$ are $O(\ell(s))$.

To turn the result for the order of magnitude into a limit theorem we need to compute the mean of the contribution of a large family. Let Y_t be the limit process defined in Lemma 6.5.

alphad **Lemma 9.3.**

$$g(x) \equiv E_x\left(\int_0^{T_0} Y_r\,dr\right) = \begin{cases} x^2/2 + x & d = 1 \\ x/(2d\beta_d) & d \ge 2. \end{cases}$$

Proof. The infinitesimal generator of Y is

$$Lf = \begin{cases} f''(x) - f'(x) & d = 1 \\ 2d\beta_d x f''(x) - 2d\beta_d x f'(x) & d \ge 2. \end{cases}$$

Intuitively, g is the solution of $Lg(x) = -x$ on $(0,\infty)$ with $g(0) = 0$, but care is needed because $(0,\infty)$ is unbounded and we have only one boundary condition. To be precise, $g = \lim_{m\to\infty} g_m$ where $Lg_m = -x$ in $(0,m)$ and $g_m(0) = g_m(m) = 0$. The limit exists since for fixed x, $m \to g_m(x)$ is increasing. From the limit result we see that g can be characterized as the minimal nonnegative solution of $Lg = -x$ on $(0,\infty)$ with $g(0) = 0$, since any other nonnegative solution has $h \ge g_m$ for all m.

In $d \ge 2$ the differential equation is

$$g''(x) - g'(x) = -\frac{1}{2d\beta_d}.$$

The solution we want is $g(x) = x/\beta_d$. To check that this is the minimal solution note that if h is another solution then $\delta = g - h$ satisfies $\delta'' - \delta' = 0$ so $\delta = c$ or $\delta(x) = e^x$. For a direct derivation of $g(x) = x/\beta_d$, note that since 0 is absorbing

$$\begin{aligned} E_x\left(\int_0^{T_0} Y_r\,dr\right) &= \int_0^\infty E_x Y_r\,dr \\ &= \int_0^\infty x e^{-2d\beta_d r}\,dr = x/(2d\beta_d) \end{aligned}$$

where the second equality follows from $(d/dr)E_xY_r = -2d\beta_d E_xY_r$.

In $d = 1$ we want to solve

$$g''(x) - g'(x) = -x$$

If we guess $g(x) = x^2/2 + x$ then $g'(x) = x + 1$ and $g''(x) = 1$ so we have a solution. Minimality holds for the same reason as before.

For a direct derivation use Itô's formula to conclude that

$$\begin{aligned} Y_t^2/2 - Y_0^2/2 &= \int_0^t Y_r\,dY_r + \frac{1}{2}\int_0^t d\langle Y\rangle_r \\ &= martingale - \int_0^t Y_r\,dr + t \end{aligned}$$

since $dY_r = \sqrt{2}\,dB_r - dr$ and $\langle Y\rangle_r = 2r$. Taking $t = T_0$, the expected value E_x, and leaving the reader to check this is legitimate, we have

$$-x^2/2 = -E_x\int_0^{T_0} Y_r\,dr + E_xT_0.$$

Since $E_xT_0 = x$ this agrees with the previous computation. □

Proof of Lemma 9.2. The number of unsuccessful type 1 mutations by time tJ/Nu_1s is $\sim tJ(1-s)/s$. Dropping the $1-s$ and using (45), the number of "large" families, i.e., those that reach ϵ/s before they die out will be $\sim tJ/\epsilon$. Using (37) we see that the expected number of man hours in a type I family after it reaches ϵ/s is

$$\frac{\ell(s)}{s}E_\epsilon\left(\int_0^{T_0} Y_r\,dr\right).$$

In Lemma 9.3 we showed the expected value $\sim \alpha_d\epsilon$ as $\epsilon \to 0$, where $\alpha_1 = 1$ and $\alpha_d = 1/(2d\beta_d)$ for $d \geq 2$. Combining our calculations and recalling that $J = 1/\ell(s)u_2$ we see that the expected number of man hours up to time tJ/Nu_1s in the families of unsuccessful type 1 mutations after they reach size ϵ/s is

$$\sim \frac{t}{\epsilon}\cdot\frac{1}{\ell(s)u_2}\cdot\frac{\ell(s)}{s}\cdot\alpha_d\epsilon = \frac{\alpha_d t}{u_2 s}. \tag{46}$$ MHL3

By repeating the of proof Lemma 2 in I, we see that the contribution from type 1 families before they reach ϵ/s (including the ones that never do), is of the same order of magnitude as (46) but with a constant that tends to 0 as $\epsilon \to 0$, so they can be ignored. The next step is to argue that if the large families evolved on independent graphical representations the result would follow from the law of large numbers. Following the same argument as in the proof of Lemma 9.1, let M denote the number of large (reach size ϵ/s) unsuccessful mutations by time tJ/Nu_1s, denote the man hours of the ith family by X_i, and the total man hours by $S_M = X_1 + \ldots + X_M$. The asymptotics of $E[S_M]$ are given by (46), and using (43) and the result of Lemma 6.4 we see that

$$Var(S_M) = O\left(\frac{t}{\epsilon \ell(s) u_2}\left(\frac{\ell(s)}{s}\right)^2 \epsilon\right) = O\left(\frac{t\ell(s)}{u_2 s^2}\right).$$

This quantity is $\ll (E[S_M])^2$ since $\ell(s) \ll 1/u_2$. Thus the fact that the total space-time volume is asymptotically $1 + o(1)$ times the mean follows from Chebyshev's inequality.

The last detail is to show that overlaps can be ignored. As discussed at the beginning of this subsection, most of the space-time volume of type 2 families that die out comes from families that reach size ϵ/s and there will be $\sim J/\epsilon$ of them by time J/Nu_1s. Lemma 6.7 implies that these families occupy a region in space time that is of size $O(\ell(s)^{d/2} \cdot \ell(s))$. If we throw J/ϵ such rectangles into a region of size $N \times J/Nu_1s$ then the probability that one of them will hit the first rectangle is of order

$$\frac{J}{\epsilon} \cdot \frac{\ell(s)^{(d+2)/2}}{J/u_1 s} = \frac{u_1 s \ell(s)^{(d+2)/2}}{\epsilon} = \rho$$

which goes to 0 by (A0).

Let $U(t)$ be the total man hours of large unsuccessful families until time $T_2(t)$ and $\tilde{U}(t)$ be the total man hours, ignoring any loss from overlaps, of large unsuccessful families until time $T_2(t)$.

From the previous calculation, $E[U(t)] \geq (1-\rho)E[\tilde{U}(t)] = (1-\rho)\alpha_d t/(u_2 s)$ where $\rho \to 0$ as $s \to 0$. Therefore $E[U(t)] \sim E[\tilde{U}(t)]$as $s \to 0$. Since $U(t) \leq \tilde{U}(t)$ due to additivity, it follows that $U(t)/\tilde{U}(t) \to 1$, and hence $U(t) \sim \alpha_d t/(u_2 s)$.

□

9.3 $K/(J+K) \to \theta \in (0,1)$

In this case the successful type 2 may come from a successsful or an unsuccessful type 1 family.

$$T_3(t) = \frac{t(J+K)}{Nu_1 s}.$$

As in the two previous cases the desired result follows immediately from

mixed_area **Lemma 9.4.** *Under the assumptions of Theorem 5 if* $K/(J+K) \to \theta \in (0,1)$ *then*

$$u_2 s \int_0^{T_3(t)} Z_1^*(r)\, dr \to \frac{\gamma_d (t/\theta)^{d+2}}{(d+1)(d+2)} \quad \textit{and} \quad u_2 s \int_0^{T_3(t)} Z_1^0(r)\, dr \to \frac{\alpha_d t}{1-\theta},$$

in probability as u_1, u_2 *and* $s \to 0$.

Proof. First from the proofs of Lemmas 9.1 and 9.2 we know that we can ignore overlaps.

We consider the Z_1^* term first. Denote the number of successful type 1 clones by $T_3(t)$ by M, and the respective man-hours by $X_1, \ldots, X_M$. Observe that $E[M] = t(J+K)$ and following the proof of Lemma 9.1 we see that conditional on $M > 0$

$$E[X_i] = \frac{\gamma_d c_d^d}{(d+1)(d+2}.$$

Therefore

$$u_2 s E[M] E[X_i] = \frac{u_2 s \gamma_d c_d^d}{(d+1)(d+2)} \frac{(t(J+K))^{d+2}}{(N u_1 s)^{d+1}} = \frac{\gamma_d}{(d+1)(d+2)} \left(\frac{t(J+K)}{K} \right)^{d+2}.$$

The result for the integral of Z_1^* then follows by analysis of the variance of $X_1 + \ldots + X_M$, which can be found in the proof of Lemma 9.1.

Next consider the Z_1^0 term. Denote the number of large (reach size ϵ/s) unsuccessful families created by $T_3(t)$ by M, and note that $E[M] \approx t(J+K)/\epsilon$. Combining this with the expected man-hours of a large family, $\frac{\ell(s)}{s}\alpha_d \epsilon$, we see that the total expected man hours time $u_2 s$ is given by

$$t\alpha_d (J+K) \ell(s) u_2 = t\alpha_d \left(\frac{J+K}{J} \right).$$

The result then follows by analysis of the variance of the number of man-hours by time $T_3(t)$ as in the proof of Lemma 9.2. □

10 Proof of Theorem 1

speedpf **Convergence to Branching Brownian Motion.** The part of the proof is from Durrett and Zähle (1997). The first step is to recall the duality between the biased voter model and coalescing random walk. For more details see Bramson and Griffeath (1981). Let η_t be the coalescing random walk in which:

(i) particles jump at rate $2d$ to a randomly chosen neighboring site.

(ii) particles give birth at rate s to a particle sent to a randomly chosen neighboring site.

(iii) if a particle lands on an occupied site (due to jump or a birth) then the two coalesce to 1

If we let η_t^B be the system starting with $\eta_0^B = B$ and let ξ_t^A be the biased voter model starting from $\xi_0^A = A$ then the two systems satisfy the duality equation:

$$P(\xi_t^A \cap B \neq \emptyset) = P(\eta_t^B \cap A \neq \emptyset).$$

In $d \geq 3$ random walks are transient, so there is positive probability β_d that an offspring will never coalesce with its parent. Durrett and Zähle (1997) show that if time is run at rate $1/s$, and space is scaled by $1/\sqrt{s}$, the coalescing random walk converges to a branching Brownian motion ζ_t in which

(i) particles perform independent Brownian motions run at rate 2,
(ii) give birth to new particles at rate β_d.

In order to achieve weak convergence they have to remove the particles that coalesce with their parents, because these result in temporary increases of the population that last (on the sped up time scale) for times of order s. To do this we ignore the new born particles for time $\tau(s) = \sqrt{s}$ before we assign them mass 1.

In $d = 2$ random walks are recurrent but the probability an offspring does not coalesce with its parent for time $> t$ is

$$\sim \pi/(\log t) \tag{47}$$ nohit

see e.g., Lemma 3.1 of Zähle, Cox and Durrett (2005). To compensate for the fact that most particles coalesce with their parents, they run time at rate $h(s) = (1/s)\log(1/s)$ and scale space by $\sqrt{h(s)}$. Furthermore we ignore the new born particles for time $\tau(s) = 1/\sqrt{\log(1/s)}$ (on the sped up time scale) before we assign them mass 1. Note that on the sped-up time scale we create new particles at rate $sh(s) = \log(1/s)$, and we assign mass to only the fraction of those that survive for $\tau(s)$ units of time, which from (47) is $\sim \pi/\log(1/s)$. Therefore we assign mass to new particles at $O(1)$ rate. Based on this the rescaled coalescing random walk converges to a branching Brownian motion ζ_t in which

(i) particles perform independent Brownian motions run at rate 2,
(ii) give birth to new particles at rate π.

At this point if one ignores the detail of interchanging two limits Theorem 2 is obvious. If particles are born at rate r and perform Brownian motions with covariance matrix $\sigma^2 I$ then the mean measure at time t for the process started with a single particle at 0 at time 0 is

$$e^{rt}\frac{1}{(2\pi t)^{d/2}}e^{-|x|^2/2\sigma^2 t} \tag{48}$$ meanmeas

Ignoring the polynomial this is 1 when

$$|x| = t\sqrt{2\sigma^2 r}$$

In our situation $\sigma^2 = 2$, $r_d = \beta_d$ in $d \geq 3$ and $r_2 = \pi$ in $d = 2$. Taking into account the space-time scaling the desired result follows.

The lower bound on the speed follows from a block construction. The argument is almost the same as in Durrett and Zähle (2007), but we have to change some details to get a sharp result. Let $I = [-1, 1]^d$, let e_1 be the first unit vector, let $v < \sqrt{4r_d}$, and for each m let $I_m = 2m(Lv)e_1 + I$. Let $\hat{\zeta}_t$ be a modification of the branching Brownian motion in which particles are killed when they land outside $[-4Lv, 4Lv]^d$. Calculations on page 1760 of Durrett and Zähle (2007) show that for any $\epsilon > 0$, we can pick L large and then K large enough so that if there are at least K particles in I_0 in $\hat{\zeta}_0$ then with probability $\geq 1-\epsilon$ we have $|\hat{\zeta}(L^2) \cap I_1| \geq K$ and $|\hat{\zeta}(L^2) \cap I_{-1}| \geq K$. Picking L large makes the mean of $|\hat{\zeta}(L^2) \cap I_1|$ large because of (48). Then picking K large controls the probability of deviations from the mean.

For integers $m \geq 0$ and n with $m + n$ even let $\theta(m, n) = 1$ if $|\hat{\zeta}(nL^2) \cap I_m| \geq K$. The result in the previous paragraph implies that η dominates 1-dependent oriented percolation with density $1 - \epsilon$. Let $r_n = \sup\{m : \theta(m, n) = 1\}$. A result in Durrett (1984), see (2) on page 1030, implies that if $\delta > 0$ and $\epsilon < \epsilon(\delta)$ then on the set where the oriented percolation does not die out, $\liminf_{n\to\infty} r_n/n \geq 1 - \delta$. This implies that for the rescaled process the edge speed is $\geq (1 - \delta)v$, which gives the lower bound.

The upper bound is proved by comparing the dual process on its original time scale with the branching process. Suppose first that $d \geq 3$. If we ignore newborn particles that will coalesce with their parents then we have a branching process in which particles are born at rate $r = \beta_d s$. We ignore coalescence events other than mother-daughter so we can project onto the x-axis to get a one-dimensional branching process Z_t at time t. The number of particles to the right of ct at time t is

$$EZ_t(ct, \infty) = e^{rt} P(S_t \geq ct) \tag{49}$$

expcheby

where S_t is a random walks that takes steps that are ± 1 with equal probability at rate 2. The steps have moment generating function

$$\phi(\theta) = \frac{e^\theta + e^{-\theta}}{2}$$

so the continuous time walk has

$$\psi_t(\theta) = E \exp(\theta S_t) = \sum_{n=0}^{\infty} e^{-2t} \frac{[2t]^n}{n!} \phi(\theta)^n = \exp\left(2t(\phi(\theta) - 1)\right).$$

Let $\delta > 0$ and let

$$a = \sqrt{\frac{2 - \delta}{1 - 2\delta} 2r} \qquad \theta_a = \frac{a}{2}.$$

As $\theta_a \to 0$, $\phi(\theta_a) - 1 \sim -\theta_a^2/2$, so if s is small

$$\phi(\theta_a) - 1 \geq -\frac{\theta_a^2}{2 - \delta}.$$

Using Markov's inequality,

$$\begin{aligned}
&P(S_t \geq (a+b)t) \leq e^{-\theta_a(a+b)t}\psi_t(\theta_a) \\
&\leq \exp\left(-t\left[\theta_a(a+b) - \frac{2\theta_a^2}{2-\delta}\right]\right) = \exp\left(-t\left[\frac{a(a+b)}{2} - \frac{a^2}{2(2-\delta)}\right]\right) \\
&= \exp\left(-t\left[\frac{ab}{2} - \frac{a^2(1-\delta)}{2(2-\delta)}\right]\right) \leq \exp\left(-t\left[\frac{1-\delta}{1-2\delta}r + \frac{ab}{2}\right]\right)
\end{aligned}$$

so using (49)

$$P(Z_t((a+b)t,\infty)) > 0) \leq \exp\left(-t\left[\frac{\delta r}{1-2\delta} + \frac{ab}{2}\right]\right).$$

To bound the spread of the biased voter model, let $\Lambda = \mathbb{Z}^d - [-at, at]^d$, By duality

$$\begin{aligned}
P(\xi_t^0 \cap [-at,at]^d \neq \emptyset) &= P(0 \in \eta_t^\Lambda) \\
&\leq \sum_{x\in\Lambda} P(0 \in \eta_t^x) \leq \sum_{k=at}^{\infty} c_d k^{d-1} P(Z_t(k,\infty) > 0) \leq e^{-ct}.
\end{aligned}$$

To prove the result in $d = 2$ we will compare the dual process with a branching random walk, where if multiple offspring land on one site they are all retained. Modifying the construction of Durrett and Zähle (2007) we ignore new born particles for time $1/s$, and add them to the dual only if they have not collided with their parents. Let Z_t denote the number of particles in the modified branching random walk by time t and let $m(t) = E[Z_t]$. From the result (47) we know that the fraction of newly created particles that are eventually added to the process is $\sim \pi/\log(1/s)$. Thus for $t > 1/s$ we have

$$m(t) = \frac{\pi s}{\log(1/s)} \int_0^{t-1/s} m(t-r)dr.$$

The previous display implies that

$$m(t) \leq \exp\left(\frac{\pi s}{\log(1/s)}t\right).$$

If we set $r = \pi s/\log(1/s)$ and repeat the calculation for the case $d \geq 3$, we obtain the desired result in $d = 2$.

References

Bach, S. et al (2000). Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* Volume 21, Issue 3, 469-476.

Beerenwinkel, N. et al. (2007) Genetic progression and the waiting time to cancer. *PLoS Comp. Bio.* Volume 3, Issue 11, 2239-2246.

Biggins, J.D. (1978) The asymptotic shape of the branching random walk. *Advance Appl. Prob.* 10, 62–84

Bozic, I., et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci.* 107, 18545–18550

Bramson, M., Cox T., Le Gall J. (2001) Super-Brownian Limits of Voter Model Clusters. *Ann. Probab.* Vol. 29 (3), pp. 1001-1032

Bramson, M., and Griffeath, D. (1980a) Asymptotics for Interacting Particle Systems on $\mathbb{Z}^d$. *Z. fur Wahr.* 53, 183-196.

Bramson, M., and Griffeath, D. (1980) On the Williams-Bjerknes tumour growth model. II. *Math. Proc. Cambridge Philos. Soc.* 88, 339–357.

Bramson, M., and Griffeath, D. (1981) On the Williams-Bjerknes tumour growth model. I. *Ann. Probab. 9*, 173–185.

Ceol, C., Pellman, D. and Zon, L. (2007). APC and colon cancer: two hits for one. *Nature Medicine.* Volume 13, 1286 - 1287.

Chai, H. and Brown, R. (2009) Field effect in cancer - an update. —it Ann. Clin. Lab. Sci. 39(4):331337.

Chatterjee, S. and Durrett, R. (2011) Asymptotic Behavior of Aldous' Gossip Process. *Ann. Appl. Probab.* 21, 2447-2482.

Cheng, H. et al. (1986) Crypt production in normal and diseased human colonic epithelium. *The Anatomical Record*, 216(1):4448.

Cox, J.T., and Perkins, E.A. (2005) Rescaled Lotka-Volterra models converge to super-Brownian motion. *Ann. Probab.* 33, 904–947.

Cox, J.T., and Perkins, E.A. (2008) Renormalization of the two-dimensional Lotka-Volterra model. *Ann. Appl. Probab.* 18, 747–812.

Cristini, V., and Lowengrub, J. (2010) *Multiscale Modeling of Cancer.* Cambridge U. Press.

Deisboeck, T.S., and Stamatakos, G.S. (2011) *Multiscale Cancer Modeling.* CRC Press.

Durrett, R. (1984) Oriented percolation in two dimensions. *Ann. Probab.* 12, 999-1040.

Durrett, R. (1995) *Ten Lectures on Particle Systems.* Pages 97-201 in *St. Flour Lecture Notes.* Lecture Notes in Math 1608. (1995). Springer-Verlag, New York.

Durrett, R., Foo, J., Leder, K., Mayberry, J., and Michor, F. (2011a) Evolutionary dynamics of tumor progression with random fitness values. Theor. Pop. Biol. 78 (2011), 54-66.

Durrett, R., Foo, J., Leder, K., Mayberry, J., and Michor, F. (2011b) Intratumor heterogeneity in evolutionary models of tumor progression. Genetics, 188 (2011), 461-477.

Durrett, R., and Mayberry, J. (2011) Traveling waves of selective sweeps. *Ann. Appl. Prob.* 21, 699–744.

Durrett, R., and Moseley, S. (2010) Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Pop. Biol.* 77, 42–48.

Durrett, R., Schmidt, D., and Schweinsberg, J. (2009) A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Prob.* 19, 676–718.

Durrett, R., and Zähle, I. (2007) On the width of hybrid zones. *Stoch. Proc. Appl.* 117, 1751–1763.

Eaden, J.A., Abrams, K.R., and Mayberry, J.F. (2001) The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut.* 48:526-535.

Ethier, S., and Kurtz, T. (1986) *Markov Processes: Characterization and Convergence.* John Wiley and Sons, New York.

Ferrari, P.L., and Prähofer, M. (2006) One-dimensional stochastic growth and Gaussian ensembles of random matrices. *Markov Processes Related Fields.* 12, 203–234.

Fisher, R.A. (1937) The wave of advance of advantageous genes. Ann. Eugenics. 7, 355–369.

Griffeath, D.S. (1978) *Additive and Cancellative Interacting Particle Systems.* Springer Lecture notes in Math 724.

Iwasa, Y., Michor, F., and Nowak, M.A. (2004) Stochastic tunnels in evolutionary dynamics. *Genetics.* 166, 1571–1579.

Iwasa, Y., Michor, F., Komarova, N.L., and Nowak, M.A. (2005) Population genetics of tumor suppressor genes. J. Theor. Biol. 233, 15–23.

Jones, S., et al. (2008) Comparative lesion sequencing provides insights into tumor evolution. *Prov. Natl. Acad. Sci.* 105, 4283–4288.

Kardar, M., Parisi, G., and Zhang, Y.C. (1986) Dynamic scaling of growing interfaces. *Phys. Rev. Letters.* 56, 889–892.

Kimura, M. (1962) On the probability of fixation of mutant genes in a population. *Genetics.* 47, 713–719.

Komarova, N.L. (2007) Spatial stochastic models of cancer initiation and progression. *Bull. Math. Biol.* 68, 1573-1599.

Komarova, N.L., Sengupta, A., and Nowak, M.A. (2003) Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J. Theor. Biol.* 223, 433–450.

Lieberman, E., Hauert, C., and Nowak, M.A. (2005) Evolutionary dynamics on graphs. *Nature.* 433, 312–316.

Luebeck, E.G. and Moolgavkar, S.H. (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *PNAS* 99 (23):1509515100.

Maley, C. et al. (2006) Genetic clonal diversity predicts progression to esophageal adeno- carcinoma. *Nature Genetics.* 38:468-473.

Martens, E.A., and Halltschek, O. (2011) Interfering waves of adaptation promote spatial mixing. *Genetics.* 189, 1045–1060.

Martens, E.A., Kostadinov, R., Maely, C.C., and Halltschek, O. (2011) Spatial structure increases the waiting time for cancer. *New J. Phys.* 13, paper 115014.

Maruyama, T. (1970) On the fixation probability of mutant genes in a subdivided population. *Genet. Res.* 15, 221–225.

Maruyama, T. (1974) A simple proof that certain quantities are independent of the geographical structure of population. *Theor. Pop. Biol.* 5, 148–154.

Merle, M. (2008) Hitting probability of a distant point for the voter model started with a single 1. *Ann. Prob.* 36, 807-861.

Merlo, L. et. al. (2006) Cancer as an evolutionary and ecological process. *Nature Reviews Cancer.* 6:924-935.

Nicolas, P. et al. (2007) The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Computational Biology* 3(3), e28.

Parzen, E. (1999) *Stochastic Processes.* Volume 24 of Classics in Applied Math. Society for Industrial and Applied Math.

Prähofer, M., and Spohn, H. (2000) Universal distributions for growth processes in 1 +1 dimensions and random matrices. *Phys. Rev. Letters.* 84, 4882–4885.

Revuz, D., and Yor, M. (1991) *Continuous Martingales and Brownian Motion.* Springer, New York.

Totafurno, J., Bjerknes, M, and Cheng, H. (1987). The Crypt Cycle: Crypt and Villus Production in the Adult Intestinal Epithelium. *Biophysical Journal.* Volume 52, Issue 2, 279294.

Williams, T., and Bjerknes, R. (1972) Stochastic model for abnormal clone spread through epithelial basal layer. *Nature.* 235, 19–21.

Zähle, I., Cox, J.T., and Durrett, R. (2005) The stepping stone model. II: Genealogies and the infinite sites model. *Annals of Applied Probability.* 15, 671-699.