The Annals of Applied Probability
2015, Vol. 25, No. 1, 104–115
DOI: 10.1214/13-AAP989

SPATIAL MORAN MODELS I. STOCHASTIC TUNNELING IN THE NEUTRAL CASE

BY RICHARD DURRETT[1] AND STEPHEN MOSELEY[2]

Duke University

We consider a multistage cancer model in which cells are arranged in a d-dimensional integer lattice. Starting with all wild-type cells, we prove results about the distribution of the first time when two neutral mutations have accumulated in some cell in dimensions $d \geq 2$, extending work done by Komarova [*Genetics* **166** (2004) 1571–1579] for $d = 1$.

1. Introduction. The accumulation of mutations is important not only for cancer initiation, progression, and metastasis, but also for the emergence of acquired resistance against chemotherapeutics, radiation therapy, or targeted drugs. For this reason there is a large and growing literature on the waiting time τ_k until some cell has acquired k prespecified mutations. In all the models we consider, type i individuals mutate to type $(i+1)$ at rate u_{i+1}. The dynamics considered have most often been studied in multi-type Moran models with a homogeneously mixing population of constant size. Here we will concentrate on how results change when one considers a spatial Moran models, and as is the case for much earlier work we will concentrate on the behavior of τ_2.

We suppose that cells of type 0 and type 1 have relative fitness 1 and λ. Since we will only consider the waiting time for the first type 2, the relative fitness of type 2's is not important. In this work we will consider situation in which λ is so close to 1 that the mutations are essentially neutral. For cancer applications, this is a restrictive assumption, and it will be removed in the companion paper (part II) by Durrett, Foo and Leder [6]. However, the current result applies to the important case of tumor suppressor genes. In that case, when both copies of the gene are inactivated trouble develops, but while there is one working copy the cell can function normally.

We begin by recalling results for the Moran model in a homogeneously mixing population of size N. Here and in what follows the mutation rates u_i and selection coefficient λ depend on N, even though this is not indicated in the notation, and we write $a_N \ll b_N$ if $a_N/b_N \to 0$ as $N \to \infty$. The next result made its first appearance on page 16,230 of Nowak et al. [17]. Since then it has appeared in

Received December 2012; revised November 2013.
[1]Supported in part by NIH Grant 5R01GM096190.
[2]This work was part of his 2011 Ph.D. thesis in Applied Mathematics at Cornell University.
MSC2010 subject classifications. 60K35, 92C50.
Key words and phrases. Biased voter model, stochastic tunneling, cancer progression.

print a number of times: [10, 11, 14, 18], and in Nowak's excellent book [16] on Evolutionary Dynamics.

THEOREM 1. *In the neutral case of the Moran model,* $\lambda = 1$, *if we assume that*

$$\frac{1}{\sqrt{u_2}} \ll N \ll \frac{1}{u_1} \tag{1}$$

and let $u_1, u_2 \to 0$ *then we have*

$$P\big(\tau_2 > t/Nu_1u_2^{1/2}\big) \to \exp(-t).$$

The same conclusion holds if $|\lambda - 1| \ll u_2^{1/2}$.

Durrett and Schmidt [7] applied these ideas to study regulatory sequence evolution and to expose flaws in Michael Behe's arguments for intelligent design. Durrett, Schmidt and Schweinsberg [8], see also Schweinsberg [19], generalized this result to cover τ_k.

The conditions in the result may look mysterious but they can be derived by simple reasoning. Here and throughout the paper and $f(u) \sim g(u)$ means $f(u)/g(u) \to 1$ as $u \to 0$. Suppose first that $\lambda = 1$.

(A1) If we start the Moran model with $k \ll N$ type 1's and the rest type 0, then the 1's behave like a critical branching process. The time needed for the 1's to die out is $O(k)$ and the number of type-1 births before they die out is $O(k^2)$. Thus we expect the first type 2 to occur in a type-1 family that reaches size $k_1 = O(1/\sqrt{u_2})$, and hence has $O(k_1^2) = O(1/u_2)$ births. The condition $1/\sqrt{u_2} \ll N$ in (1) guarantees $k_1 \ll N$.

(A2) Since the voter model is a martingale, the probability a type-1 mutation creates a family that reaches size $1/\sqrt{u_2}$ is $\sqrt{u_2}$. More to the point a simple computation (consider what happens at the first jump) shows that the probability a type-1 family gives rise to a type 2 before it dies out is $\sim \sqrt{u_2}$. Since mutations to type 1 occur at times of a rate Nu_1 Poisson process and with probability $\sim \sqrt{u_2}$ give rise to a type 2, it follows that if ρ_2 is the birth time of the type-1 mutant that first gives rise to a type 2 then

$$P\big(\rho_2 > t/Nu_1u_2^{1/2}\big) \to \exp(-t).$$

To complete the proof we need to show that $\tau_2 - \rho_2 \ll \rho_2$, and for this we need the condition $N \ll 1/u_1$ in (1).

(A3) By the discussion of (A1), the first mutation will occur in a family that reaches a size $O(1/\sqrt{u_2})$. If $|\lambda - 1| \ll u_2^{1/2}$, then computations with Girsanov's formula show that (in the limit $u_2 \to 0$) the behavior of the Moran model, while it is $O(1/\sqrt{u_2})$, is indistinguishable from the case with no drift.

The assumption of a homogeneously mixing cell populations simplifies calculations considerably, but is not realistic for solid tumors. For this reason, Komarova [12] considered a spatial model, which is very similar to one introduced much earlier by Williams and Bjerknes [20]. Due to work of Bramson and Griffeath [2, 3], the second model is known to probabilists as the biased voter model.

In the usual formulation of the biased voter model, each site on the d-dimensional integer lattice $\mathbb{Z}^d$ can be in state 0 or 1 indicating the presence of a cell with relative fitness 1 or $\lambda > 1$. Cells give birth at a rate equal to their fitness, and the offspring replaces a nearest neighbor chosen at random. When $\lambda = 1$ this is the voter model which was introduced independently by Clifford and Sudbury [4] and Holley and Liggett [9]. For a summary of what is known see Liggett [15].

In the biased voter model births drive the process. In Komarova's version cells die at rate 1 and are replaced by a copy of a nearest neighbor chosen with probability proportional to its fitness. A site with n_i neighbors in state i makes

transitions	at rate
$0 \to 1$	$\lambda n_1/(\lambda n_1 + n_0)$
$1 \to 0$	$n_0/(\lambda n_1 + n_0)$

In $d = 1$ if the set of sites in state 1 is an interval $[\ell, r]$ with $\ell < r$ then any site that can change has $n_1 = n_0 = 1$ so Komarova's model is a time change of the biased voter model. In $d \geq 2$ this is not exactly true. However, we are interested in values of $\lambda = 1 + s$ where $s = 0.02$ or even less, so we expect the two models to have very similar behavior. In any case, the difference between the two models is much less than their difference from reality, so we will choose to study the biased voter, whose duality with branching coalescing random walk (to be described below) gives us a powerful tool for doing computations.

Since we want a finite cell population we will restrict our process to be a subset of $(-L/2, L/2]^d$. Komarova [12] uses "Dirichlet boundary conditions", that is, she assumes her space is an interval with no cells outside, but this is awkward because the set of type-1 cells may reach one end of the interval and then no further changes happen at that end. To avoid this, we will use periodic boundary conditions, that is, we consider $(\mathbb{Z} \bmod L)^d$. The resulting toroidal geometry is a little strange for studying cancer. However, using $(\mathbb{Z} \bmod L)^d$ has the advantage that the space looks the same seen from any point. Our results will show that for the parameter values the first type 2 will arise when the radius of the set of sites occupied by 1's is $\ll L$ so the boundary conditions do not matter.

Let ξ_s^0 be the set of cells equal to 1 in the voter model with no mutations from 0 to 1 on $\mathbb{Z}^d$ starting from a single type 1 at 0. Let $|\xi_s^0|$ be the number of cells in ξ_s^0, and let

$$\nu_d = 1 - E\exp\left(-u_2 \int_0^{T_0} |\xi_s^0|\, ds\right). \tag{2}$$

This quantity, which is defined for the voter model without mutation, calculates the probability, which depends on the dimension d, that a mutation to type 1 gives

rise to a type 2 before its family dies out. To see why this is true note that the integral $\int_0^{T_0} |\xi_s^0|\,ds$ gives the total number of man-hours in the type-1 family, and conditional on this the number of mutations that will occur is Poisson with mean $u_2 \int_0^{T_0} |\xi_s^0|\,ds$.

Since mutations to type 1 in a population of N cells occur at rate Nu_1 this suggests that

$$P(\tau_2 > t) \to \exp(-Nu_1 v_d t). \tag{3}$$

As we will explain in a moment, there is a constant γ_d so that $v_d \sim \gamma_d h_d(u_2)$ as $u_2 \to 0$ where

$$h_d(u) = \begin{cases} u^{1/3}, & d = 1, \\ u^{1/2}\log^{1/2}(1/u), & d = 2, \\ u^{1/2}, & d \geq 3. \end{cases} \tag{4}$$

To state the result we need one more definition:

$$g_d(u) = \begin{cases} u^{1/3}, & d = 1, \\ \log^{-1/2}(1/u), & d = 2, \\ 1, & d \geq 3. \end{cases} \tag{5}$$

THEOREM 2. *In the neutral case of the biased voter model, $\lambda = 1$, if we assume*

$$\frac{1}{h_d(u_2)} \ll N \ll \frac{g_d(u_2)}{u_1}, \tag{6}$$

then there are constants γ_d given in (12) and (13) so that as $u_1, u_2 \to 0$

$$P\big(\tau_2 > t/Nu_1\gamma_d h_d(u_2)\big) \to \exp(-t).$$

The same conclusion holds if $|\lambda - 1| \ll h_d(u_2)$.

In $d = 1$ this result was proved by Komarova [12], see her equation (62) and assumption (60), then change notation $u_1 \to u$, $u_2 \to u_1$. See also her survey paper [13]. Note that when $d \geq 3$ the order of magnitude of the waiting time and the assumptions are the same as in Theorem 1. In $d = 2$ there are logarithmic corrections to the behavior in Theorem 1, so only in the case of $d = 1$ (which is relevant to cancer in the mammary ducts) does space make a substantial change in the waiting time.

The reasons for the conditions in Theorem 2 are the same as in Theorem 1.

(B1) We will see that the mutation to type 2 will occur in a type-1 family that reaches size $k = O(1/h_d(u_2))$. The left-hand side assumption in (6) implies that $k \ll N$, so the type-2 mutant arises before the 1's reach fixation.

(B2) Let ρ_2 be the time of the first type-1 mutation that begins the family that eventually leads to a type 2. Since mutations to type 1 occur at rate Nu_1 and lead to a type 2 with probability ν_d, it is easy to see that

$$P(\rho_2 > t) \to \exp(-Nu_1\nu_d t)$$

so to prove the result we need to show that with high probability $\tau_2 - \rho_2 \ll \rho_2$. As the reader will see, this is guaranteed by the right-hand side assumption in (6).

(B3) As in the discussion of Theorem 1, once we know that the mutation to type 2 will occur in a type-1 family that reaches size $k = O(1/h_d(u_2))$, it follows that if $|\lambda - 1| \ll h_d(u_2)$ then (in the limit $u_2 \to 0$) the behavior of the size of the biased voter $|\xi_t^0|$ is, while it is $O(1/h_d(u_2))$, indistinguishable from the case with no drift.

2. The key to the proof. The size of the voter model, when $|\xi_t^0| > 0$, is a time change of symmetric simple random walk, with jumps happening at two times the size of the boundary $|\partial\xi_t^0|$, which is the number of nearest neighbor pairs with $x \in \xi_t^0$ and $y \notin \xi_t^0$. The one-dimensional case is easy because when $\xi_t^0 \neq \varnothing$ the boundary $|\partial\xi_t^0| = 2$. The key to the study of the process in $d \geq 2$ is the observation that there are constants β_d so that

$$|\partial\xi_t^0| \sim_p \begin{cases} 2d\beta_d|\xi_t^0|, & d \geq 3, \\ 4\beta_2|\xi_t^0|/\log(|\xi_t^0|), & d = 2, \end{cases} \tag{7}$$

where $|\partial\xi_t^0| \sim_p f(|\xi_t^0|)$ means that when $|\xi_t^0|$ is large, $|\partial\xi_t^0|/f(|\xi_t^0|)$ is close to 1 with high probability.

The intuition behind this result is that the voter model is dual to a collection of coalescing random walks, so in $d \geq 3$ neighbors of points in ξ_t^0 will be unoccupied with probability $\approx \beta_d$, the probability two simple random walks started at 0 and $e_1 = (1, 0, \ldots, 0)$ never hit. In dimension $d = 2$, the recurrence of random walks implies that when $|\xi_t^0| = k$ is large, most neighbors of points in ξ_t^0 will be occupied, but due to the fat tail of the recurrence time sites will be vacant with probability $\sim \beta_2/\log k$, where $\beta_2 = \pi$.

Before we try to explain why (7) is true, we will list an important consequence. Let T_k be the first time $|\xi_t^0| = k$. Let

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

LEMMA 1. *Let ξ_t^0 be the unbiased voter model (i.e., $\lambda = 1$) starting from a single occupied site.*

$$\left(\frac{|\xi^0_{T_{n\varepsilon}+a_n t}|}{n} \middle| T_{n\varepsilon} < \infty \right) \Rightarrow (Y_t | Y_0 = \varepsilon), \tag{8}$$

where $\Rightarrow$ indicates convergence in distribution of the stochastic processes and the limit has

$$dY_t = \begin{cases} \sqrt{2}\, dB_t, & d = 1, \\ \sqrt{2\beta_d Y_t}\, dB_t, & d \geq 2, \end{cases}$$

where B_t is a one-dimensional Brownian motion. In $d = 1$ the process is stopped when it hits 0. In $d \geq 2$, 0 is an absorbing boundary so we do not need to stop the process.

In $d = 1$ this result is trivial. If one accepts (7) then (8) can be proved easily by computing infinitesimal means and variances and using standard weak convergence results. In $d \geq 2$, (7) and (8) are almost consequences of work of Cox, Durrett and Perkins [5]. They speed up time at rate a_n, scale space by $1/\sqrt{a_n}$, and assign each point occupied in the voter model mass $1/n$ to define a measure-valued diffusion X^n which they prove converges to super-Brownian motion. See their Theorem 1.2. (Their scaling is a little different in $d = 2$ but this makes no difference to the limit.)

Let $V'_{n,s}(x)$ be the fraction of sites adjacent to x in state 0 at time s (with the prime indicating that we multiply this by $\log n$ in $d = 2$, see page 196). A key step in the proof in [5] is to show, see (I1) on page 202, that for nice test functions ϕ

$$E\left[\left(\int_0^T X_s^n(\{V'_{n,s} - \beta_d\}\phi^2)\, ds\right)^2\right] \to 0, \tag{9}$$

where $X_s^n(f)$ denote the integral of the function f against the measure X_s^n. The result in (9) shows that when we integrate in time and average in space (multiplying by a test function to localize the average) then (7) is true.

From the convergence of the measure valued diffusion X^n to super-Brownian motion, (8) follows by considering the total mass. Earlier we said (8) is almost a consequence of [5], since they start their process from an initial measure [i.e., $O(n)$ initial 1's] while consider a single occupied site and condition on reaching $n\varepsilon$. However, this defect can be remedied by citing the work of Bramson, Cox and LeGall [1], who have a result, Theorem 4 on page 1012 that implies (8) in $d \geq 2$.

The result in (8) is enough for Section 3, but for the calculations in Section 4 we will need a version of (7). In that section we will compute under the assumption that if $|\xi_t^0| = k$

$$|\partial \xi_t^0| = \begin{cases} 2d\beta_d k, & d \geq 3, \\ 4\beta_2 k/\log k, & d = 2. \end{cases} \tag{10}$$

If one wants to give a rigorous proof of the estimates there, then small values of k, can be treated with the inequalities

$$Ck^{1/d} \leq |\partial \xi_t^0| \leq 2dk,$$

and one can control large values of k using (9) and estimates such as (J1) and (J2) on page 208 of [5]. We will assume (10) in order to avoid getting bogged down in technicalities.

3. Proof, part I. Let v_d^ε be the probability defined in (2) ignoring mutations to type 2 that occur before $T_{n\varepsilon}$. The size of the voter model, $|\xi_t^0|$, is a martingale, so if we let P_1 to denote the law of the voter model starting from one occupied site $P_1(T_{n\varepsilon} < \infty) = 1/n\varepsilon$. Applying (8) now,

$$v_d^\varepsilon \sim \frac{1}{n\varepsilon} \cdot \left[1 - E_\varepsilon \exp\left(-na_n u_2 \int_0^{T_0} Y_s\, ds\right)\right], \tag{11}$$

where $T_0 = \min\{t : Y_t = 0\}$, E_ε is the expected value for $(Y_t | Y_0 = \varepsilon)$. We have

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

So if we set $n = 1/h_d(u_2)$ then (4) implies $na_n u_2 \to 1$ and using (11) gives

$$v_d^\varepsilon \sim h_d(u_2) \cdot \left[\frac{1 - E_\varepsilon \exp(-\int_0^{T_0} Y_s\, ds)}{\varepsilon}\right].$$

Thus the type-2 mutation will occur in a family that reaches sizes $O(1/h_d(u_2))$, and we must assume $1/h_d(u_2) \ll N$.

If we ignore the time to reach size $1/h_d(u_2)$, the time needed to generate the type-2 mutation is, by (8), of order

$$a\big(1/h_d(u_2)\big) \sim \begin{cases} u_2^{-2/3}, & d = 1, \\ 2u_2^{-1/2} \log^{1/2}(1/u_2), & d = 2, \\ u_2^{-1/2}, & d \geq 3, \end{cases}$$

where we have written $a(n)$ for a_n for readability. Thus for (B2) we need $a(1/h_d(u_2)) \ll 1/Nu_1 h_d(u_2)$, which means $N \ll g_d(u_2)/u_1$.

The next order of business is to compute v_d. Stochastic calculus (or calculations with infinitesimal generators) tells us that

$$v(x) = E_x \exp\left(-\int_0^{T_0} Y_s\, ds\right)$$

is the unique function on $[0, \infty)$ with values in $[0, 1]$, $v(0) = 1$ and

$$v'' - xv = 0 \qquad \text{in } d = 1, \qquad \beta_d x v'' - xv = 0 \qquad \text{in } d \geq 2.$$

In $d = 1$ all solutions have the form:

$$v(x) = \alpha Ai(x) + \beta Bi(x),$$

where Ai and Bi are Airy functions

$$Ai(x) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{t^3}{3} + xt\right) dt,$$

$$Bi(x) = \frac{1}{\pi} \int_0^\infty \exp\left(-\frac{t^3}{3} + xt\right) + \sin\left(\frac{t^3}{3} + xt\right) dt.$$

Since Bi is unbounded and Ai is decreasing on $[0,\infty)$, we take $\beta = 0$ and set $\alpha = 3^{2/3}\Gamma(2/3)$ to satisfy the boundary condition, $v(0) = 1$. Letting $\varepsilon \to 0$ we conclude that

$$\gamma_1 = -\alpha Ai'(0) = 3^{1/3}\Gamma(2/3)/\Gamma(1/3). \tag{12}$$

In $d \geq 2$, $v(x) = \exp(-\beta_d^{-1/2}x)$, and we have

$$\gamma_d = \beta_d^{-1/2}. \tag{13}$$

4. Proof, part II: Missing details for $\lambda = 1$. In the previous section we have calculated the probability ν_d^ε that a type-1 family reaches size $\varepsilon/h_d(u_2)$ and then gives rise to a type 2. To let $\varepsilon \to 0$ and prove Theorem 2 we need to consider the possibility of a mutation to type 2 in a family that (i) never reaches size $n\varepsilon$, or (ii) will reach $n\varepsilon$ but has not yet. To have a convenient name we will call these small families. Families of the first kind arise at rate $Nu_1(1 - 1/n\varepsilon)$ and families of the second kind arise at rate $Nu_1/n\varepsilon$. We will now calculate the expected rate at which type 2's are born from these small families. In the proof of Theorem 2, we will let $\varepsilon \to 0$ slowly as $n \to \infty$ so we can and will assume $n\varepsilon \to \infty$.

Consider the voter model ξ_t^0 starting from a single 1 at the origin at time 0. Let V_k be the total time spent at level k, that is, $|\{t : |\xi_t^0| = k\}|$ and let N_k be the total number of returns to level k before leaving the interval $(0, n\varepsilon)$. Recalling our assumption in (10), we let $q(k)$ the rate jumps occur at level k.

Let S_k be the embedded discrete time chain, which is a simple random walk, and let $T_k^+ = \min\{n \geq 1 : S_n = k\}$.

$$\begin{aligned} E_1\left(\int_0^{T_0} |\xi_s^0|\, ds \,\middle|\, T_0 < T_{n\varepsilon}\right) &= E_1\left(\sum_{k=1}^{n\varepsilon} kV_k \,\middle|\, T_0 < T_{n\varepsilon}\right) \\ &= E_1\left(\sum_{k=1}^{n\varepsilon} \frac{kN_k}{q(k)} \,\middle|\, T_0 < T_{n\varepsilon}\right) \\ &= \sum_{k=1}^{n\varepsilon} \frac{\overline{P}_1(T_k < \infty)}{\overline{P}_k(T_k^+ > T_0)} \frac{k}{q(k)}, \end{aligned} \tag{14}$$

where the bar indicates conditioning on $T_0 < T_{n\varepsilon}$. A similar argument shows that

$$E_1\left(\int_0^{T_{n\varepsilon}} |\xi_s^0|\, ds \,\middle|\, T_{n\varepsilon} < T_0\right) = \sum_{k=1}^{n\varepsilon} \frac{1}{\widehat{P}_k(T_k^+ > T_{n\varepsilon})} \frac{k}{q(k)}, \tag{15}$$

where the hat indicates conditioning on $T_{n\varepsilon} < T_0$.

The three conditional probabilities we need can be computed using facts about simple random walk that follow from the fact that it is a martingale.

$$\overline{P}_1(T_k < \infty) = \frac{P_1(T_k < \infty)P_k(T_0 < T_{n\varepsilon})}{P_1(T_0 < T_{n\varepsilon})} = \frac{(1/k)(1 - k/n\varepsilon)}{(1 - 1/n\varepsilon)}. \tag{16}$$

For the next two we note that the first step has to be in the correct direction for these events to happen.

$$\overline{P}_k(T_k^+ > T_0) = \frac{(1/2)(1/k)}{(1-k/n\varepsilon)}, \tag{17}$$

$$\widehat{P}_k(T_k^+ > T_{n\varepsilon}) = \frac{(1/2)(1/(n\varepsilon-k))}{(k/n\varepsilon)}. \tag{18}$$

Thus the expected total man-hours $\int_0^{T_0} |\xi_s^0|\,ds$ for a family that will die out before reaching size $n\varepsilon$ is

$$\sim \frac{2}{(1-1/n\varepsilon)} \sum_{k=1}^{n\varepsilon} (1-k/n\varepsilon)^2 \frac{k}{q(k)}, \tag{19}$$

and in families that have yet to reach size $n\varepsilon$,

$$\frac{2}{n\varepsilon} \sum_{k=1}^{n\varepsilon} (n\varepsilon - k) \frac{k^2}{q(k)}. \tag{20}$$

The next result shows that the contribution of small families are indeed negligible. Note that in all three cases the order of magnitude of the contributions from small families is the same as the overall rate, but contains a constant that $\to 0$ as $\varepsilon \to 0$.

LEMMA 2. *The expected total man-hours in small families is*

$$\leq \begin{cases} Nu_1u_2^{1/3} \cdot \dfrac{\varepsilon^2}{4}, & d=1, \\ Nu_1u_2^{1/2} \log^{1/2}(1/u_2) \cdot \dfrac{7\varepsilon}{24\beta_2}, & d=2, \\ Nu_1u_2^{1/2} \cdot \dfrac{\varepsilon}{2d\beta_d}, & d \geq 3. \end{cases}$$

PROOF. In one dimension, $q(k) = 2$. The sum in (19) is dominated by

$$\int_0^{n\varepsilon} (1-x/n\varepsilon)^2 x\,dx = \frac{1}{(n\varepsilon)^2} \int_0^{n\varepsilon} y^2(n\varepsilon - y)\,dy = \frac{(n\varepsilon)^2}{12}.$$

Thus, families of the first kind produce type 2's at rate $\leq Nu_1u_2(n\varepsilon)^2/12$. The expression in (20) is dominated by

$$\frac{2}{n\varepsilon} \int_0^{n\varepsilon} (n\varepsilon - x)x^2\,dx = \frac{(n\varepsilon)^3}{6}.$$

Thus, families of the second kind produce type 2's at rate $\leq Nu_1u_2(n\varepsilon)^2/6$. Adding the last two conclusions gives the result for $d = 1$.

In $d \geq 3$, (10) implies $q(k) = 2d\beta_d k$, so (19) becomes

$$\frac{1}{d\beta_d} \sum_{k=1}^{n\varepsilon} (1 - k/n\varepsilon)^2.$$

The sum is bounded above by the integral

$$\int_0^{n\varepsilon} (1 - x/n\varepsilon)^2 \, dx = \frac{n\varepsilon}{3},$$

so with our choice of $n = u_2^{-1/2}$, families of the first kind produce type 2's at rate bounded above by $Nu_1 u_2^{1/2}\varepsilon/(3d\beta_d)$. Setting $q(k) = 2d\beta_d k$, (20) becomes

$$\frac{1}{d\beta_d n\varepsilon} \sum_{k=1}^{n\varepsilon} (n\varepsilon - k)k.$$

The sum is bounded above by the integral

$$\int_0^{n\varepsilon} (n\varepsilon - x)x \, dx = \frac{(n\varepsilon)^3}{6}.$$

Thus, families of the second kind produce type 2's at rate $\leq Nu_1 u_2^{1/2}\varepsilon/(6d\beta_d)$. Adding the last two conclusions gives the result for $d \geq 3$.

In $d = 2$, (10) implies $q(k) = 4\beta_2 k/\log k$, so (19) becomes

$$\frac{1}{2\beta_2} \sum_{k=1}^{n\varepsilon} (1 - k/n\varepsilon)^2 \log k.$$

Each term in the sum is bounded above by $\log(n\varepsilon)$, so the sum is less than $n\varepsilon \log n\varepsilon$. Since $n = u_2^{-1/2} \log^{-1/2}(1/u_2)$, families of the first kind produce type 2's at rate bounded above by

$$Nu_1 u_2 \cdot \frac{1}{2\beta_2} n\varepsilon \log(n\varepsilon) = Nu_1 u_2 \cdot \frac{1}{2\beta_2} \varepsilon u_2^{-1/2} \log^{-1/2}(1/u_2) \cdot \frac{1}{2} \log(1/u_2)$$
$$= \frac{\varepsilon}{4\beta_2} Nu_1 u_2^{1/2} \log^{1/2}(1/u_2).$$

Taking $q(k) = 4\beta_2 k/\log k$, (20) becomes

$$\frac{1}{2\beta_2 n\varepsilon} \left(\sum_{k=1}^{n\varepsilon} (n\varepsilon - k)k \log k \right).$$

The sum is bounded above by

$$\int_0^{n\varepsilon} (n\varepsilon - x)x \log(n\varepsilon) \, dx \leq \frac{(n\varepsilon)^3}{6} \log(n\varepsilon).$$

Thus families of the second kind produce type 2's at rate bounded above by

$$\frac{Nu_1u_2}{n\varepsilon}\cdot\frac{1}{2\beta_2 n\varepsilon}\cdot\frac{(n\varepsilon)^3}{6}\log(n\varepsilon)=\frac{1}{12\beta_2}Nu_1u_2\cdot n\varepsilon\log(n\varepsilon)$$
$$=\frac{\varepsilon}{24\beta_2}Nu_1u_2^{1/2}\log^{1/2}(1/u_2).$$

Adding the last two conclusions gives the result for $d=2$ and completes the proof. □

5. Proof, part III: Almost neutral mutations. In the biased voter model, whose law we denote by P^λ, jumps occur at rate $1+\lambda$ times the size of the boundary. To compensate for this we need to run the unbiased ($\lambda=1$) voter at rate $(1+\lambda)/2$. If we do this, call the resulting law $\widetilde{P}^0$, and let ω_T is a realization of ξ_t^0 run up to time T then the Radon–Nikodym derivative

$$\frac{dP^\lambda}{d\widetilde{P}^0}(\omega_T)=\left(\frac{2\lambda}{\lambda+1}\right)^{n_+}\left(\frac{2}{\lambda+1}\right)^{n_-},$$

where n_+ and n_- are the number of up jumps in ω_t when $0\le t\le T$.

If $\max_{t\le T}|\xi_t^0|=O(K)$ then the difference $0\le n_+-n_-=O(K)$. Since under $\widetilde{P}_0$, $|\xi_t^0|$ is a time change of simple random walk, we see that the total number of jumps $n_++n_-=O(K^2)$. Taking $K=1/h_d(u_2)$ and assuming $|\lambda-1|\ll h_d(u_2)$, when u_2 is small the Radon–Nikodym derivative is

$$=\left(1+\frac{\lambda-1}{\lambda+1}\right)^{n_+}\left(1-\frac{\lambda-1}{\lambda+1}\right)^{n_-}$$
$$=\left(1+\frac{\lambda-1}{\lambda+1}\right)^{n_+-n_-}\left(1-\frac{(\lambda-1)^2}{(\lambda+1)^2}\right)^{n_++n_-}\approx 1.$$

The last result implies that (8) extends to almost neutral mutations, and the computations in Section 2 are valid. To extend the part of the proof in Section 3, we need to check that (16)–(18) are true asymptotically for almost neutral mutations. To do this we recall that if $a<x<b$

$$P_x^\lambda(T_b<T_a)=\frac{\theta^x-\theta^a}{\theta^b-\theta^a}\qquad\text{where }\theta=1/\lambda. \tag{21}$$

When $0\le a<x\le b=O(1/h_d(u_2))$ and $|\lambda-1|\ll h_d(u_2)$ we have

$$P_x^\lambda(T_b<T_a)\approx\frac{x-a}{b-a}.$$

To show that the sums come out the same we need the following uniform version which follows from (21). If $|\lambda-1|h_d(u_2)\to 0$ then for any C fixed

$$\sup_{0\le -a,b\le C/h_d(u_2)}\left|\frac{P_0^\lambda(T_b<T_a)}{-a/(b-a)}-1\right|\to 0.$$

REFERENCES

[1] BRAMSON, M., COX, J. T. and LE GALL, J.-F. (2001). Super-Brownian limits of voter model clusters. *Ann. Probab.* **29** 1001–1032. MR1872733

[2] BRAMSON, M. and GRIFFEATH, D. (1980). On the Williams–Bjerknes tumour growth model. II. *Math. Proc. Cambridge Philos. Soc.* **88** 339–357. MR0578279

[3] BRAMSON, M. and GRIFFEATH, D. (1981). On the Williams–Bjerknes tumour growth model. I. *Ann. Probab.* **9** 173–185. MR0606980

[4] CLIFFORD, P. and SUDBURY, A. (1973). A model for spatial conflict. *Biometrika* **60** 581–588. MR0343950

[5] COX, J. T., DURRETT, R. and PERKINS, E. A. (2000). Rescaled voter models converge to super-Brownian motion. *Ann. Probab.* **28** 185–234. MR1756003

[6] DURRETT, R., FOO, J. and LEDER, K. (2013). Spatial Moran models, II. Tumor growth and progression.

[7] DURRETT, R. and SCHMIDT, D. (2008). Waiting for two mutations: With applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics* **180** 1501–1509.

[8] DURRETT, R., SCHMIDT, D. and SCHWEINSBERG, J. (2009). A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Probab.* **19** 676–718. MR2521885

[9] HOLLEY, R. A. and LIGGETT, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.* **3** 643–663. MR0402985

[10] IWASA, Y., MICHOR, F., KOMAROVA, N. L. and NOWAK, M. A. (2005). Population genetics of tumor suppressor genes. *J. Theoret. Biol.* **233** 15–23. MR2122451

[11] IWASA, Y., MICHOR, F. and NOWAK, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics* **166** 1571–1579.

[12] KOMAROVA, N. L. (2006). Spatial stochastic models for cancer initiation and progression. *Bull. Math. Biol.* **68** 1573–1599. MR2257717

[13] KOMAROVA, N. L. (2007). Loss- and gain-of-function mutations in cancer: Mass-action, spatial and hierarchical models. *J. Stat. Phys.* **128** 413–446. MR2331192

[14] KOMAROVA, N. L., SENGUPTA, A. and NOWAK, M. A. (2003). Mutation–selection networks of cancer initiation: Tumor suppressor genes and chromosomal instability. *J. Theoret. Biol.* **223** 433–450. MR2067856

[15] LIGGETT, T. M. (1999). *Stochastic Interacting Systems*: *Contact*, *Voter and Exclusion Processes*. Springer, Berlin. MR1717346

[16] NOWAK, M. A. (2006). *Evolutionary Dynamics*: *Exploring the Equations of Life*. Belknap Press, Cambridge, MA. MR2252879

[17] NOWAK, M. A., KOMAROVA, N. L., SENGUPTA, A., JALLEPALLI, P. V., SHIH, I. M., VOGELSTEIN, B. and LENGAUER, C. (2002). The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci.* **99** 16226–16231.

[18] NOWAK, M. A., MICHOR, F., KOMAROVA, N. L. and IWASA, Y. (2004). Evolutionary dynamics of tumor suppressor gene inactivation. *Proc. Natl. Acad. Sci. USA* **101** 10635–10638.

[19] SCHWEINSBERG, J. (2008). The waiting time for m mutations. *Electron. J. Probab.* **13** 1442–1478. MR2438813

[20] WILLIAMS, T. and BJERKNES, R. (1972). Stochastic model for abnormal clone spread through epithelial basal layer. *Nature* **235** 19–21.

DEPARTMENT OF MATHEMATICS
DUKE UNIVERSITY
BOX 90320
DURHAM, NORTH CAROLINA 27708-0320
USA
E-MAIL: rtd@math.duke.edu