Spatial Moran model

Rick Durrett

...

A report on joint work
with Jasmine Foo, Kevin Leder (Minnesota),
and Marc Ryser (postdoc at Duke)

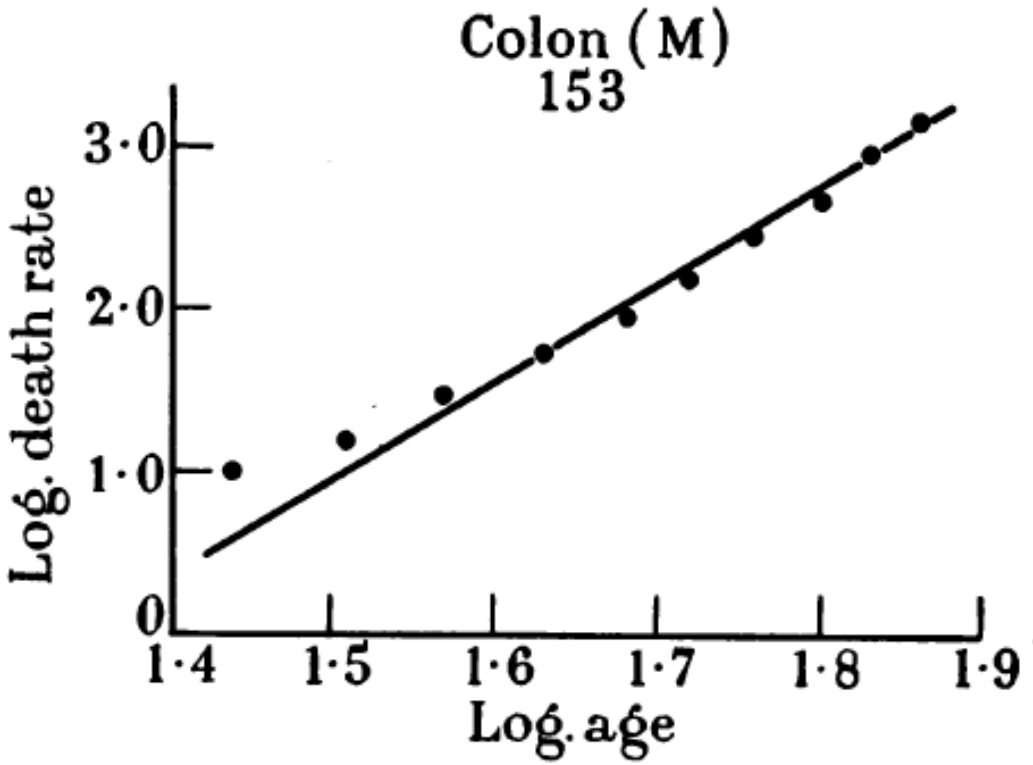
Stan Ulam once said:

"I have sunk so low that my last paper contained numbers with decimal points."

Figure: Feynman, Ulam, and von Neumann

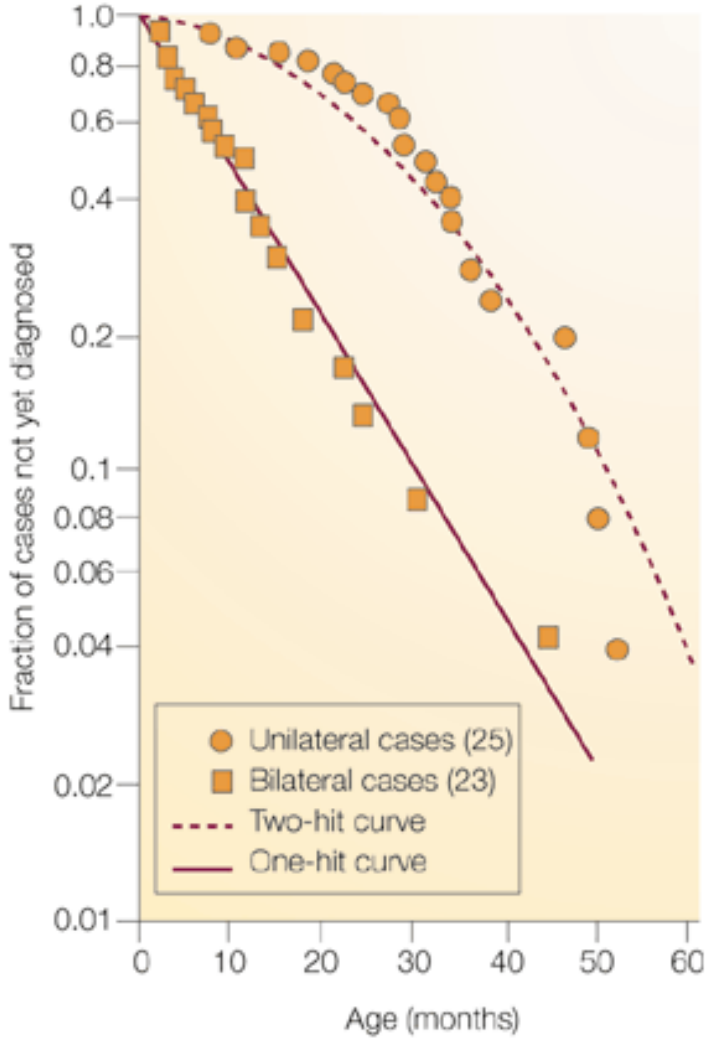
Armitage and Doll (1954)

Noticed that log-log plots of cancer incidence data are linear for a large number of cancer types; for example, colorectal cancer incidence has a slope of 5.18 in men and 4.97 in women. Concluded based on calculating the distribution of the sum of expoonentials that the number of stages = slope + 1.



Incidence of Retinoblastoma

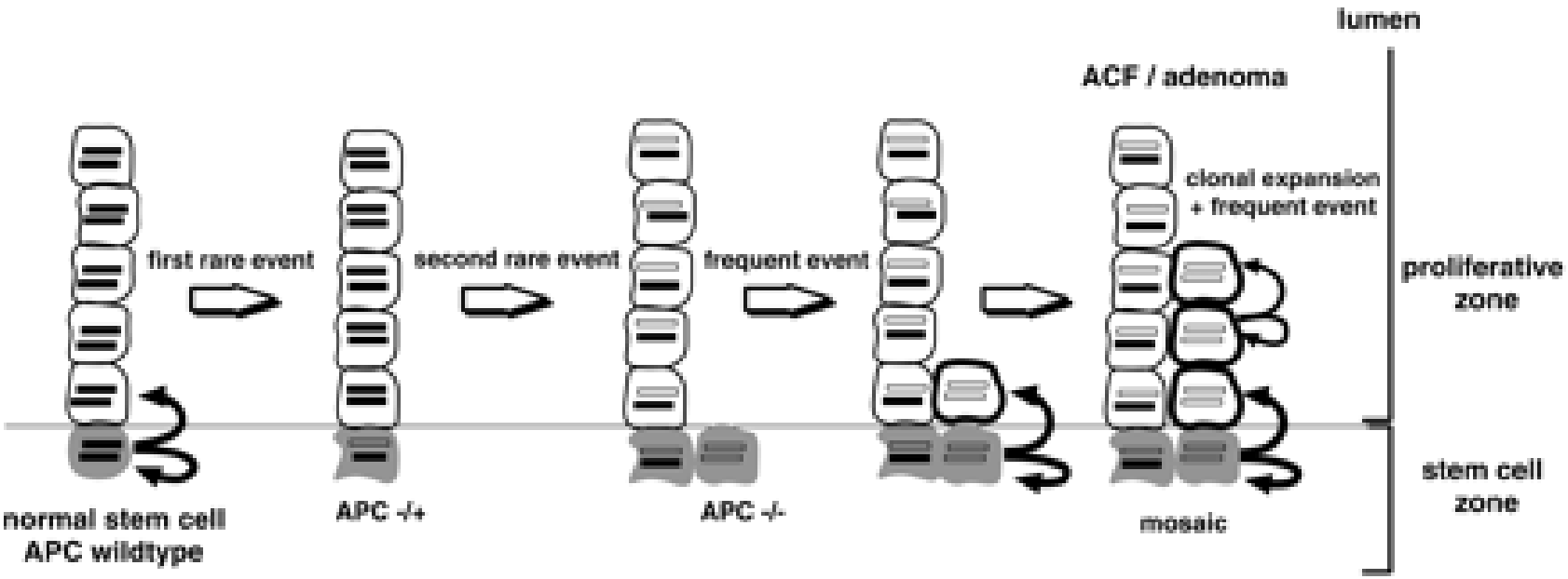
Knudson's two hit hypothesis $\longrightarrow$ tumor-suppressor genes



Nature Reviews | Cancer

Multi-stage theory of carcinogenesis

Luebeck and Moolgavakar (2002) PNAS fit a four stage model to incidence of colon cancer by age.



What are the stages ?

- In sporadic cases of colon cancer the first two stages are inactivation of the tumor suppressor gene *APC* adenomatous polyposis coli.
- KRAS is an **oncogene** (one mutation turns it on). Once it is turned on it recruits and activates proteins necessary for the propagation of growth factor
- The final stage is thought to involve the inactivation of *TP*53 the gene which makes *p*53.

The real situation is much less clear cut: in some cases *APC* is not knocked out but the oncogene β-catenin is upregulated.

Motivating Problem: Field Cancerization

This term originated in the 1950s from observations of Slaughter et al on HNSCC: cells in the tissue around a malignancy show premalignant transformations. This effect resulting a higher than expected prevalence of multiple local second primary tumors and the presence of synchronous distant tumors has been observed in a number of other cancers.

Our goal is to use a simple spatial cancer model to make quantitative predictions about how the geometry of the field depends on underlying parameters such as selective advantage and mutation rate, and ultimately provide guidance on surgical excision margins.

Modeling Philosophy

We will study a very simple model that can be analyzed mathematically instead of using a more complex model that can only be analyzed by simulation, e.g., "hybrid discrete models" where the evolution of the cells is coupled to PDE for one of more nutrients. See e.g., the book by Cristintini and Lowengrub (2010) or publications by Sandy Anderson et al.

Although some people have strong preferences for one style over the other, the two approaches afford complimentary insights.

Spatial Moran model

Williams and Bjerknes (1972)

Takes place on $(\mathbb{Z} \bmod L)^d$. $N = L^d$

Type i individuals mutate to type $(i+1)$ at rate u_{i+1}.

Cells of type i have relative fitness $\prod_{j=1}^{i}(1+s_j)$.

Cells give birth at rate equal to their fitness and replace one of their nearest neighbors chosen at random.

In homogeneously mixing case of Moran model, we replace one of the N individuals chosen at random.

Almost Neutral case

Suppose $s = 0$ and consider the homogeneously mixing case Let τ_2 be the time of birth of the first type 2 individual, and write $a_N \ll b_N$ if $a_N/b_N \to 0$.

The next result comes from work done by Komarova, Iwasa, Michor, and Nowak in various combinations during 2002–2005.

Stochastic Tunneling. *If* $1/\sqrt{u_2} \ll N \ll 1/u_1$ *then*

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

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

Type 2 mutation occurs in a type 1 family that reaches size $1/u_2^{1/2}$, which is $\ll N$.

Spatial result

	$d=1$	$d=2$	$d \geq 3$
$h_d(u)$	$u^{1/3}$	$u^{1/2}\log^{1/2}(1/u)$	$u^{1/2}$
$g_d(u)$	$u^{1/3}$	$\log^{-1/2}(1/u)$	1

Theorem. *If* $1/h_d(u_2) \ll N \ll g_d(u_2)/u_1$ *then*

$$P(\tau_2 > t/Nu_1 h_d(u_2)) \to \exp(-\alpha_d t)$$

The same conclusion holds if $|s| \ll h_d(u_2)$.

Type 2 mutation occurs in a type 1 family that reaches size $1/h_d(u_2)$.

$d = 1$ Komarova (2007). $d \geq 2$. Durrett and Moseley (2011) AoAP 2014.

$d \geq 3$ is like homogeneously mixing, logarithmic corrections in $d = 2$.

Concrete Example

Cells have diameter ≈ 10 microns or $10^{-5} m$.

$N = 10^8$, $(10cm)^2$

$u_1 = 10^{-8}$ fairly specific nonsynonymous mutation

$s_1 = 10^{-2}$

$u_2 = 10^{-6}$ gene knockout

$s_2 = 0.04$

In most real applications there will be more stages.

Head and Neck Squamous Cell Carcinoma

Type	fitness
Normal	1
Atypia	1
Dysplasia	$1 + s_1$
Carcinoma in situ	$(1 + s_1)(1 + s_2)$
Invasive	

Work in progress. Rather than base model on poorly understood genetic events, we look at phenotypic changes. To derive waiting times combine results for stochastic tunneling with results with selection.

Spatial Moran is time change of random walk

Suppose only two types (0 and 1) and no mutation. $s_1 = s$.

On each boundary edge connecting a 1 with a 0, there is a competition with the interactions along this edge 0 changing to 1 at rate $1 + s$ and the 1 changing to 0 at rate 1.

Let $\xi_t = \{x : \eta_t(x) = 1\}$. While $\xi_t \neq \emptyset$, the size of the set, $|\xi_t|$, is a time change of an asymmetric simple random walk which makes jumps $+1$ with $p = (1 + s)/(2 + s)$ and and -1 with probability $1 - p = 1/(2 + s)$.

$$P_1(T_0 = \infty) = 1 - \frac{1-p}{p} = \frac{s}{1+s} \sim s \quad \text{as } s \to 0.$$

Bramson and Griffeath (1980,1981)

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

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

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

Speed

Let e_1 be the first unit vector and define $c_d(s)$ so that the intersection of D with the x axis is $[-c_d(s)e_1, c_d(s)e_1]$.

The proof of Bramson and Griffeath implies that $c_d(s) \geq b_d s$ where b_d is a positive constant. Using ideas of Durrett and Zähle (2007):

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

$$c_d(s) \sim \begin{cases} s & d = 1 \\ \sqrt{4\pi s/\log(1/s)} = 0.1651 & d = 2 \\ \sqrt{4\beta_d s} & d \geq 3, \end{cases}$$

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.

σ_1 time of the first successful type 1

Consider the process on the torus $(\mathbb{Z} \bmod L)^d$ and let $N = L^d$.

"Obvious" Fact. *If* $s_1, u_1 \to 0$ *then*

$$P(\sigma_1 > t/Nu_1s_1) \to e^{-t}$$

In concrete example $Nu_1s_1 = 100$ cell divisions

"Proof" While the fraction of the torus covered by 1's is small, mutations to type 1 occur at rate Nu_1 and are successful with probability $\sim s_1$.

Problem successive attempts are not independent.

σ_1 continued

A long lasting unsuccessful type 1 mutations will survive a time of order

$$\ell(s) = \begin{cases} s^{-2} & d = 1, \\ s^{-1}\log(1/s) = 460.5 & d = 2, \\ s^{-1} & d \geq 3. \end{cases} \tag{2}$$

and cover a volume in space with diameter $O(\ell(s)^{1/2})$

Theorem 2. *Suppose (A0)* $(1/u_1) \gg \ell(s_1)^{(d+2)/2}$. *As* $s_1, u_1 \to 0$

$$P(\sigma_1 > t/Nu_1s_1) \to e^{-t}$$

$u_1\ell(s_1)^{(d+2)/2} \ll 1$ means we can ignore two type 1 mutations in one space time box of volume $\ell(s_1)^{d/2} \cdot \ell(s_1)$.

In concrete example (A0) is $10^8 \gg 2.11 \times 10^5$

An uninteresting regime

By Theorem 2, the time until the first successful type 1 mutation will be

$$t_{mut} = \Theta(1/L^d u_1 s_1).$$

Since successful mutations spread at rate $c_d(s_1)$, the time for a successful mutation to spread to cover the torus will be

$$t_{fix} = \Theta(L/c_d(s_1))$$

When we have

$$(SF) \qquad L \ll L_c = \left(\frac{c_d(s_1)}{s_1 u_1}\right)^{1/(d+1)} = 1181 \text{ versus } L = 10^4$$

then we will have *sequential fixation*: successful mutations will fix faster than they arise. When (SF) holds, the times between successive mutations will be exponential with mean $1/t_{mut}$.

Takeover by type 1's

A site x will be type 1 at time t if there is a successful type 1 mutation in the space-time cone $\{(y,r) : 0 \le r \le t, |y-x| < c_d(t-r)\}$. Such mutations are approximately a Poisson process with rate $u_1 s$ so

$$P(x \in \xi_t) \approx 1 - \exp\left(-u_1 s_1 \int_0^t \gamma_d (c_d r)^d \, dr\right) \approx 1 - \exp\left(-u_1 s_1 \frac{\gamma_d c_d^d t^{d+1}}{d+1}\right)$$

where γ_d is volume of the unit ball in d dimensions. $P(x \in \xi_t)$ goes from density ϵ to $1-\epsilon$ at times of order

$$(1/s_1 u_1 c_d^d)^{1/(d+1)} = L_c/c_d = 7153.$$

This observation is useful to show that with high probability the first successful type 2 mutation will occur while the density of 1's is small.

Mea culpa

In order to study σ_2, we consider a simplified model in which:

(i) successful type 1 mutations give rise to a deterministic linearly growing ball;

(ii) we ignore the effect of unsuccessful type 1 and type 2 mutations on the growth in (i)

(iii) we flip coins to see if an unsuccessful type 1 will give rise to a successful type 2.

Even with these simplifications it takes 20 pages to write the proofs.

Three scenarios for σ_2.

1. The successful type 2 mutation arises in the first successful type 1 family, and before the time of the second successful type 1 mutation.

2. There are several successful type 1 mutations before the first successful type 2 mutation.

3. The number of successful type 1 mutations before the first successful type 2 mutation tends to ∞. Not relevant for cancer.

In both cases 1 and 2 the malignancy is surrounded by premalignant cells but in scenario 2 there is also a "distant field."

A mysterious but important constant

A cone with slant $c_d(s_1)$ and height $1/Nu_1s_1$ has volume

$$V = \int_0^{1/Nu_1s} \gamma_d (c_d r)^d \, dr = \frac{\gamma_d}{d+1} c_d^d \left(\frac{1}{Nu_1 s_1} \right)$$

$\Gamma V = 1/u_2 s_2$ when

$$\Gamma = (Nu_1 s_1)^{d+1} (c_d^d(s_1) u_2 s_2)^{-1} = 9.1716$$

Γ is the number of cones of this height needed to generate a successful type 2 mutation.

If $\Gamma \to 0$ then the successful type 2 mutation will occur in the first successful type 1 family.

Condition (A1)

If we let $t_2 = (c_d^d u_2 s_2)^{-1/(d+2)}$ then

$$\int_0^{t_2} (d+1)(c_d r)^d \, dr = 1/u_2 s_2$$

At time t_2 the radius of the clone is

$$c_d t_2 = (c_d/u_2 s)^{1/(d+1)} = 66.7 \text{ versus } L = 1000$$

To fit in the torus we need

$$(A1) \qquad (c_d/u_2 s)^{d/(d+1)} \ll L^d = N$$

Condition (A2)

If $\bar{\xi}_t$ is the biased voter model conditioned to die out

$$E_1\left(\int_0^\infty |\bar{\xi}_t|\, dt\right) \sim C_d \ell(s_1)$$

Expected number of unsuccessful type 1's before the first successful type 1 is $1/s$ so the probability that an unsuccessful type 1 family gives rise to a successful type 2 mutation, before there are J successful type 1 mutations is

$$\frac{J}{s_1} \cdot \ell(s_1) \cdot u_2 s_2$$

To rule this out we want

$$(A2) \qquad u_2 s_2 \frac{\ell(s_1)}{s_1} \ll 1$$

In concrete example LHS is 1.84×10^{-3}

Scenarios 1 and 2

Assume (A0), (A1), (A2).

Theorem 3. *If* $\Gamma \to 0$ *then* $P(\sigma_2 > t/Nu_1s_1) \to \exp(-t)$.

Theorem 4. If $\Gamma \to I \in (0, \infty)$ then

$$P(\sigma_2 > t/Nu_1s_1) \to \exp\left(-\int_0^t 1 - \exp\left[-\frac{\gamma_d}{I} \cdot \frac{y^{d+1}}{d+1}\right] dy\right)$$

If $I = 0$ the second result reduces to the first. Since $1 - e^{-x} \le x$ the integral is $\le C_d y^{d+2}/I$ and so the number of successful type 1 mutations needed to generate a successful type 2 is of order $K = \Gamma^{1/d+2}$

Note that for large t, RHS $\approx e^{-t}$.

Foo, Leder, and Ryser

Use these results to compute the distribution of:

the area of the local field (the ball in which the mutation occurs) at time σ_2;

the number of patches of type 1's and the area of the distant field;

the time until a second primary tumor in the distant field.