Probability models for cancer development and progression

Rick Durrett

John Mayberry (Cornell)
Stephen Moseley (Cornell)
Deena Schmidt (MBI)
Jason Schweinsberg (UCSD)

Armitage and Doll (1954)
log-log plots of incidence versus age

Why a power law?

If mutations from stage \( i - 1 \) to stage \( i \) occur at rate \( u_i \), the probability density of reaching stage \( k \) at time \( t \) is

\[
\approx u_1 u_2 \cdots u_k t^{k-1} \left( \frac{1}{(k-1)!} \right)
\]

so the slope is the number of stages \(-1\).

Slopes Stomach: 5.91 M, 5.27 F; Pancreas M 5.76, F 6.48

Incidence of Retinoblastoma

Knudson’s two hit hypothesis

Progression to Colon Cancer

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

The Problem

Given a population of size \( N \), how long does it take until \( \tau_k \) the first time we have an individual with a prespecified sequence of \( k \) mutations?

- Initially all individuals are type 0.
- Each individual is subject to replacement at rate 1.
- A copy is made of an individual chosen at random from the population.
- Type \( j - 1 \) mutates to type \( j \) at rate \( u_j \).
Theorem. If \( Nu_1 \to 0 \) and \( N \sqrt{u_2} \to \infty \)

\[
P(t_2 > t/Nu_1 \sqrt{u_2}) \to e^{-t}
\]

10,000 simulations of \( n = 10^3, u_1 = 10^{-4}, \sqrt{u_2} = 10^{-2} \)

**Waiting for \( k \) mutations**

Total progeny of a critical binary branching process has

\[
P(\xi > k) \sim C k^{-1/2},
\]

so the sum of \( M \) such random variables is \( O(M^2) \).

To get 1 individual of type 4, we need of order

\[
1/u_4 \text{ births of type 3.}
\]

\[
1/\sqrt{u_2} \text{ mutations to type 3.}
\]

\[
1/u_3 \sqrt{u_4} \text{ births of type 2.}
\]

\[
1/u_2 u_3^{1/2} u_4^{1/4} \text{ mutations to type 2.}
\]

\[
1/u_2 u_3^{1/2} u_4^{1/4} \text{ births of type 1.}
\]

\[
1/u_2 u_3^{1/4} u_4^{1/8} \text{ mutations to type 1.}
\]

**Idea of Proof**

Since 1’s mutate to 2’s at rate \( u_2 \), \( \tau_2 \) will occur when there have been \( O(1/u_2) \) births of individuals of type 1.

The number of 1’s is roughly a (time change of ) symmetric random walk, so \( \tau_2 \) will occur when the number of 1’s reaches \( O(1/\sqrt{u_2}) \).

\( N \gg 1/\sqrt{u_2} \) guarantees that up to \( \tau_2 \) the number of 1’s is \( o(N) \), so 1 mutations occur at rate \( Nu_1 \), and 1’s that have 2 descendants occur at rate \( Nu_1 \sqrt{u_2} \)

The waiting time from the 1 mutation until the 2 mutant appears is of order \( 1/\sqrt{u_2} \). For this to be much smaller than the overall waiting time \( 1/Nu_1 \sqrt{u_2} \) we need \( Nu_1 \ll 1 \).

**Small time behavior**

Most cancers occur in less than 1% of the population so we are looking at the lower tail of the distribution. Let \( g_k(t) = Q_k(t) \leq t \) where \( Q_k \) is the probability for the branching process started with one type 1. In the case \( u_j \equiv \mu \)

\[
g'_k(t) = \mu g_{k-1}(t) - (1 - \mu) g_k(t)^2 - 2 \mu g_k(t)
\]

One can inductively solve the differential equations and finds

If \( t \ll \mu^{-1/2} \) then \( g_k(t) \approx \mu^{-1} t^{k-1} / (k-1)! \)


**Durrett, Schmidt, and Schweinsberg**

Probability type \( j \) has a type \( k \) descendant.

\[
\sim t_{j,k} = u_{j+1}^{1/2} u_{j+2}^{1/4} \cdots u_k^{1/2-j} \text{ for } 1 \leq j < k
\]

**Theorem.** Let \( k \geq 2 \). Suppose that:

(i) \( Nu_1 \to 0 \).

(ii) For \( j = 1, \ldots, k-1, u_{j+1}/u_j \geq b_j \) for all \( N \).

(iii) There is an \( a > 0 \) so that \( N^a u_k \to 0 \).

(iv) \( Nu_{1,k} \to \infty \).

Then for all \( t > 0 \), \( \lim_{N \to \infty} P(\tau_k > t/Nu_1 \tau_{1,k}) = \exp(-t) \).

**Exponentially growing population, 1**

Joint work with Stephen Moseley.

Some chronic myeloid leukemia patients show resistance to imatinib at diagnosis, and many others develop resistance during the first year of treatment. Iwasa, Nowak, and Michor (2006) and Haeno, Iwasa, Michor (2007), both in Genetics.

Model is a multi-type branching process in which type \( i \) cells have \( i \geq 0 \) mutations.

Type \( i \) cells give birth at rate \( a_i \) and die at rate \( b_i \).

\( \lambda_i = a_i - b_i \) increases in \( i \).

Type \( i \)'s mutate at rate \( u_{i+1} \) becoming type \( i+1 \).
Math questions

Compute the distribution of $\tau_k$ be the time of the occurrence of the first type $k$. $k = 1, 2$ most relevant to development of immunity.

Let $Z_k(t)$ be the number of type $k$ cells at time $t$. Find the limiting behavior of $e^{-\lambda_k t}Z_k(t)$.

$$P(\tau_1 > t| Z_0(s), s \leq t, \Omega_\infty^0) = \exp \left( -u_1 \int_0^t Z_0(s) ds \right)$$

$$(e^{-\lambda_k t} Z_k(s) | \Omega_\infty^0) \sim V_0 = \text{exponential}(\lambda_0 / a_0)$$

$$P(\tau_1 > t| \Omega_\infty^0) \approx E \exp \left( -u_1 V_0 e^{\lambda_1 t} / \lambda_0 \right) = \frac{\lambda_0}{\lambda_0 + a_0 u_1 e^{\lambda_1 t} / \lambda_0}$$

Growth of the 1's

$(Z_1^0, Z_2^0, \ldots, Z_k^0)$ is a decomposable Galton-Watson process, which Kesten and Stigum studied in discrete time.

For any $k \geq 1$

$$M_k^t = e^{-\lambda_k t} Z_k(t) - \int_0^t u_k e^{-\lambda_k s} Z_{k-1}(s) ds$$

is a martingale

Show that $M_k^t$ is $L^2$ bounded and conclude

Theorem. $e^{-\lambda_1 t} Z_1(t) \rightarrow W_1$ a.s. with

$$EW_1 = u_1 / (\lambda_1 - \lambda_0).$$

$W_1$ has a power law tail

Let $Z_i^*(t)$ be the number of type-$i$'s at time $t$ in a system with $Z_i^0(t) = e^{\lambda_i t} V_0$ for all $t \in (-\infty, \infty)$.

Theorem. $e^{-\lambda_i t} Z_i^*(t) \rightarrow V_i$ a.s. with

$$Ee^{-\theta V_i} = 1 / (1 + u_1 c_{V_i} \theta^{ \lambda_0 / \lambda_i})$$

and hence

$$P(V_i > x) \sim c_{V_i} x^{-\lambda_0 / \lambda_i}$$

Actually $W_1$ does not have a power law tail but $P(W_1 \neq V_1) = u_1 a_0 / \lambda_0^2$ is small.

Proof of power law tail

Results from Sjoblom et al (2006): 35 tumors

Results from Wood et al. (2007) Science
Flawed Methodology?

Wood et al. (2007): 40 of the top 119 genes, selected based on the pathways in which they occur, were chosen for further sequencing. 15 of the 40 genes (38%) were not mutated in any of the 96 tumors studied.
False Discovery Rate of 10 % ??

Exponentially growing population, 2

Joint work with John Mayberry
In some cancers, e.g., colon cancer, an early stage is the growth of a “benign tumor,” before progression to malignancy.
Genetic Progression and the Waiting Time to Cancer
Niko Beerenwinkel, Tibor Antal, David Dingli, Arne Traulsen, Kenneth W. Kinzler, Victor E. Velculescu, Bert Vogelstein, Martin A. Nowak
PLoS Computational Biology 3 (2007) e225
Wright-Fisher model in exponentially growing population. Cells with $k$ mutations have relative fitness $(1 + \gamma)^k$.  

Figure: Simulation from Beerenwinkel et al.
Second regime $\alpha \in (3/2, 11/6)$. Limit for $\alpha = 1.82$

Third regime $\alpha \in (11/6, 25/12)$. Limit for $\alpha = 1.95$

Exponentially growing population

Figure: Simulation from Beerenwinkel et al.