A waiting time problem arising from the study of multi-stage carcinogenesis

by Jason Schweinsberg

University of California at San Diego

(partly joint work with Rick Durrett and Deena Schmidt)

Outline of talk

1. History and background
2. The model
3. Main results and proof sketches
4. Generalizations
5. Open problems

History and Background

Muller (1951): "There are, however, reasons for inferring that many or most cancerous growths would require a series of mutations in order for cells to depart sufficiently from the normal."

Armitage-Doll (1954): Proposed multi-stage model of cancer. If a cell has experienced $j-1$ changes, jth change at rate u_j. Cancer occurs after m changes.

- For small t, probability that $m-1$ changes happen before time t is approximately

$$\frac{u_1 u_2 \dots u_{m-1} t^{m-1}}{(m-1)!}.$$

- Probability that mth change happens in $[t, t+dt]$ is $u_m\, dt$.
- Incidence rate of cancer at time t, for small t, is

$$r(t)\, dt \approx \frac{u_1 u_2 \dots u_m t^{m-1}}{(m-1)!}\, dt.$$

Examined data on 17 types of cancer. Typically incidence rate increases like 5th or 6th power of age, suggesting 6 or 7 stages.

Knudson (1971) discovered that retinoblastoma is a result of two mutations.

Moolgavkar-Luebeck (1992, 2002): 3 or 4 stage model for colon cancer.

Calabrese et. al. (2005): between 4 and 9 mutations required for colon cancer.

Sjoblom et. al. (2006): as many as 14 mutations involved in colon cancer, 20 in breast cancer.

Moolgavkar-Luebeck (1992): "the concept of multi-stage carcinogenesis is one of the central dogmas of cancer research."

Regulatory sequence evolution (Durrett-Schmidt, 2007): DNA sequences (6-9 nucleotides) control how genes are expressed. Several mutations required to get a given regulatory sequence.

The Model

Population has fixed size N.

Moran model: each individual lives for an Exponential(1) time, then gets replaced by individual chosen at random.

Individuals experience mutations at rate μ (depends on N).

Let τ_m be the first time at which an individual has m mutations.

Clearly $\tau_1 \sim$ Exponential($N\mu$).

Problem: For $m \geq 2$, find limiting distribution of τ_m as $N \to \infty$.

Important simplifying assumptions:

- One mutation rate μ, rather than jth mutation at rate u_j.
- Mutations offer no selective advantage.

Two ways to accumulate m mutations

1. Fixation: one mutation spreads to the entire population. Then we have to wait for $m-1$ additional mutations.

2. Stochastic tunneling (Iwasa-Michor-Nowak, 2004): one individual gets m mutations before any mutation fixates.

$4m-3$ different regimes for asymptotic distribution of τ_m.

For $m=2$, results for 4 of 5 regimes in Iwasa-Michor-Komarova-Nowak (2005), Wodarz-Komarova (2005).

We have complete results for all m, with rigorous proofs.

Focus in this talk on $m=3$.

Preliminaries: critical branching processes

When a mutation occurs, number of individuals with the mutation evolves as follows:

- $k \to k-1$ at rate $k(N-k)/N$.
- $k \to k+1$ at rate $k(N-k)/N$.

When $k \ll N$, approximate by a continuous-time branching process, each individual gives birth and dies at rate 1.

P(number with the mutation reaches L) $= 1/L$.

P(mutation lasts for time t) $\sim C/t$ (Kolmogorov, 1938).

Preliminaries: multitype branching processes

Consider the following multitype branching process:

- Initially there is a single type 1 individual.
- Every individual gives birth and dies at rate 1.
- Type k individual mutates to type $k+1$ at rate μ.

Let $p_m = P(\text{a type } m \text{ individual is born eventually})$. Then

$$p_m = \frac{1}{2+\mu}(0) + \frac{1}{2+\mu}(2p_m - p_m^2) + \frac{\mu}{2+\mu}p_{m-1}.$$

Can rewrite as $p_m^2 + \mu p_m - \mu p_{m-1} = 0$, so

$$p_m = \frac{-\mu + \sqrt{\mu^2 + 4\mu p_{m-1}}}{2} \approx \sqrt{\mu p_{m-1}}.$$

Since $p_1 = 1$, solving inductively gives

$$p_m \approx \mu^{1-2^{-(m-1)}}.$$

Exponential and Gamma limits for small μ

By time t, there are approximately $N\mu t$ mutations, most successful lasts for a time $O(N\mu t)$ before disappearing or fixating.

If $N\mu \ll 1$, time between when a mutation occurs and when it disappears or fixates is much smaller than t, can be neglected.

Mutation fixates with probability $1/N$, has descendant with m mutations with probability approximately p_m. Fixation happens first if $1/N \gg p_m$, or

$$N\mu^{1-2^{-(m-1)}} \to 0.$$

Fixation before 2 mutations if $\mu \ll N^{-2}$, before 3 if $\mu \ll N^{-4/3}$.

Theorem: Let Z_1 and Z_2 be independent Exponential(1).

- If $\mu \ll N^{-2}$, then $\mu\tau_3 \to_d Z_1 + Z_2$.
- If $N^{-2} \ll \mu \ll N^{-4/3}$, then $\mu\tau_3 \to_d Z_1$.
- If $N^{-4/3} \ll \mu \ll N^{-1}$, then $N\mu^{7/4}\tau_3 \to_d Z_1$.

Remarks about proof

Expected time for mutation to disappear or fixate is $O(\log N)$.

Expected time, up to time t, that there is a mutation in the population that has not fixated is $O(N\mu t \cdot \log N)$.

If $\mu \ll 1/(N \log N)$, ignore first mutations that occur while there is another mutation in the population, proofs are easy.

When $C/(N \log N) \leq \mu \ll 1/N$, many mutations in population at once. Need to show the events that they have a descendant with three mutations are approximately independent.

Poisson Approximation

Split $[0, t]$ into M intervals, let A_i be event that mutation in ith interval has descendant with 3 mutations.

Let β_i be all intervals within a distance $C\mu^{-3/4}$ of the ith.

Result below gives exponential waiting time.

Lemma (Arratia-Goldstein-Gordon, 1989): Let W be number of the A_i that occur, $\lambda = E[W]$. Let $\mathcal{F}_i = \sigma((A_j)_{j \notin \beta_i})$. Define

$$b_1 = \sum_{i=1}^{M} \sum_{j \in \beta_i} P(A_i) P(A_j),$$

$$b_2 = \sum_{i=1}^{M} \sum_{i \neq j \in \beta_i} P(A_i \cap A_j),$$

$$b_3 = \sum_{i=1}^{M} E\Big[\big| P(A_i | \mathcal{F}_i) - P(A_i) \big| \Big].$$

Then $|P(W = 0) - e^{-\lambda}| \leq b_1 + b_2 + b_3$.

The borderline cases

Suppose $\mu \sim CN^{-2}$. After the first fixation, each mutation fixates with probability $1/N$ and has a descendant with 3 mutations but does not fixate with probability $O(1/N)$.

Let $X(t)$ be the number of individuals with the mutation at time t. Births and deaths at rate $X(t)(N - X(t))/N$. Additional mutations happen at rate $\mu X(t)$.

Consider instead a simple random walk $(Y(t), t \geq 0)$ which jumps at rate 1. Mutation rate

$$\mu Y(t) \cdot \frac{N}{2Y(t)(N - Y(t))} = \frac{\mu}{2(1 - Y(t)/N)}.$$

Probability of no fixation or additional mutation is

$$E\left[\exp\left(-\frac{\mu}{2} \int_0^T \frac{1}{1 - Y(t)/N}\, dt \right) \mathbf{1}_{\{Y(T)=0\}} \right],$$

where T is the first time the walk hits 0 or N.

If $Y(0) = \lfloor Nx \rfloor$, limit as $N \to \infty$ is

$$u(x) = E\left[\exp\left(-\frac{C}{2} \int_0^U \frac{1}{1 - B(t)}\, dt \right) 1_{\{B(U)=0\}} \right],$$

where $(B(t), t \geq 0)$ is Brownian motion with $B(0) = x$ and U is the first time Brownian motion hits 0 or 1.

Use Feynman-Kac to get differential equation for $u(x)$, obtain series solution, calculate

$$\alpha = \lim_{x \to 0} \frac{1 - u(x)}{x} = \sum_{k=1}^{\infty} \frac{C^k}{(k-1)!(k-1)!} \Big/ \sum_{k=1}^{\infty} \frac{C^k}{k!(k-1)!}.$$

Theorem: Let $Z \sim$ Exponential(1), $Y \sim$ Exponential(α).

- If $\mu \sim CN^{-2}$, then $\mu\tau_3 \to_d Z + Y$.
- If $\mu \sim CN^{-4/3}$, then $\mu\tau_3 \to_d Y$ (use $C^{3/2}$ in definition of α).

The case $N\mu \nrightarrow 0$

Limit not exponential because we can't ignore the time between the first mutation and the third mutation.

Let $X_k(t)$ be number of individuals with k mutations at time t.

$$E[X_1(t)] \approx N\mu t, \qquad E[X_2(t)] \approx \mu \int_0^t E[X_1(s)]\, ds \approx \frac{N\mu^2 t^2}{2}.$$

Fluctuations primarily from births and deaths, so

$$\text{Var}(X_1(t)) = O(N\mu t^2), \qquad \text{Var}(X_2(t)) = O(N\mu^2 t^3).$$

$X_1(t) \approx E[X_1(t)]$ when $\sqrt{N\mu t^2} \ll N\mu t$, or $N\mu \gg 1$.

$X_2(t) \approx E[X_2(t)]$ when $\sqrt{N\mu^2 t^3} \ll N\mu^2 t^2$, or $N\mu^2 t \gg 1$.

If $N\mu^2 t \gg 1$, we have $X_2(t) \approx E[X_2(t)] \approx N\mu^2 t^2/2$, so

$$P(\tau_3 > t) \approx \exp\left(-\mu \int_0^t E[X_2(s)]\, ds \right) \approx \exp\left(-\frac{N\mu^3 t^3}{6} \right).$$

This is relevant when $N\mu^2(N^{-1/3}\mu^{-1}) \gg 1$, or $\mu \gg N^{-2/3}$.

When $N^{-1} \ll \mu \ll N^{-2/3}$, we have $X_1(t) \approx E[X_1(t)] \approx N\mu t$. Second mutations are in the population for only a short time, so

$$P(\tau_3 > t) \approx \exp\left(-\mu p_2 \int_0^t E[X_1(s)]\, ds \right) \approx \exp\left(-\frac{N\mu^{5/2} t^2}{2} \right).$$

Theorem:

- If $\mu \gg N^{-2/3}$, then

$$\lim_{N\to\infty} P(N^{1/3}\mu\tau_3 > t) = \exp(-t^3/6).$$

- If $N^{-1} \ll \mu \ll N^{-2/3}$, then

$$\lim_{N\to\infty} P(N^{1/2}\mu^{5/4}\tau_3 > t) = \exp(-t^2/2).$$

Two more borderline cases

When $\mu \sim CN^{-1}$, get stochastic effects both from the number of individuals with one mutation, and from the time between the first and third mutations.

When $\mu \sim CN^{-2/3}$, get stochastic effects both from the number of individuals with two mutations, and from the time between the second and third mutations.

Consider the following two-type branching process:

- Initially there is a single type 1 individual.
- Every individual gives birth and dies at rate 1.
- A type 1 individual mutates to type 2 at rate r.

Let $f(r,t) = P(\text{a type 2 individual appears by time } t)$.

Solving Kolmogorov's backward equations gives (for small r),

$$f(r,t) \approx \sqrt{r} \cdot \frac{1 - e^{-2\sqrt{r}t}}{1 + e^{-2\sqrt{r}t}}.$$

When $\mu \sim CN^{-1}$, mutations happen at rate $N\mu$. Mutation at time s has probability approximately $f(\mu p_2, t-s)$ of having a descendant with 3 mutations by time t, so

$$P(\tau_3 > t) \approx \exp\left(-\int_0^t N\mu f(\mu p_2, t-s)\, ds \right).$$

When $\mu \sim CN^{-2/3}$, there are $N\mu s$ individuals with one mutation at time s. Second mutations happen at rate $N\mu^2 s$, so

$$P(\tau_3 > t) \approx \exp\left(-\int_0^t N\mu^2 s f(\mu, t-s)\, ds \right).$$

Theorem:

- If $\mu \sim CN^{-1}$, then

$$\lim_{N\to\infty} P(\mu^{3/4}\tau_3 > t) = \exp\left(-C\int_0^t \frac{1-e^{-2(t-s)}}{1+e^{-2(t-s)}}\, ds \right).$$

- If $\mu \sim CN^{-2/3}$, then

$$\lim_{N\to\infty} P(\mu^{1/2}\tau_3 > t) = \exp\left(-C^{3/2}\int_0^t \frac{s(1-e^{-2(t-s)})}{1+e^{-2(t-s)}}\, ds \right).$$

Results for general m

Theorem: Let $S_j \sim \text{Gamma}(j,1)$ and $Y_j \sim \text{Exponential}(\alpha_j)$,

$$\alpha_j = \sum_{k=1}^{\infty} \frac{C^{2k(1-2^{-(j-1)})}}{(k-1)!(k-1)!} \Big/ \sum_{k=1}^{\infty} \frac{C^{2k(1-2^{-(j-1)})}}{k!(k-1)!}.$$

- If $\mu \ll N^{-2}$, then $\mu\tau_m \to_d S_{m-1}$.
- If $N^{-2^{j-1}/(2^{j-1}-1)} \ll \mu \ll N^{-2^j/(2^j-1)}$ for $j = 2, \ldots, m-1$, then $\mu\tau_m \to_d S_{m-j}$.
- If $N^{-2^{m-1}/(2^{m-1}-1)} \ll \mu \ll N^{-1}$, then $N\mu^{2-2^{-(m-1)}}\tau_m \to_d S_1$.
- If $\mu \sim CN^{-2^{j-1}/(2^{j-1}-1)}$ for some $j = 2, \ldots, m$ and $C > 0$, then $\mu\tau_m \to_d S_{m-j} + Y_j$.

- If $\mu \gg N^{-2/m}$, then

$$\lim_{N\to\infty} P(\tau_m > N^{-1/m}\mu^{-1}t) = \exp\left(-\frac{t^m}{m!}\right).$$

- If $N^{-1/(1+(m-j-2)2^{-(j+1)})} \ll \mu \ll N^{-1/(1+(m-j-1)2^{-j})}$ for some $j = 1, \ldots, m-2$, then

$$\lim_{N\to\infty} P(\tau_m > N^{-1/(m-j)}\mu^{-1-(1-2^{-j})/(m-j)}t) = \exp\left(-\frac{t^{m-j}}{(m-j)!}\right).$$

- If $\mu \sim CN^{-1/(1+(m-j-1)2^{-j})}$ for some $j = 1, \ldots, m-1$, then

$$\lim_{N\to\infty} P(\tau_m > \mu^{-(1-2^{-j})}t)$$
$$= \exp\left(-\frac{C^{1+(m-j-1)2^{-j}}}{(m-j-1)!}\int_0^t \frac{s^{m-j-1}(1-e^{-2(t-s)})}{1+e^{-2(t-s)}}\,ds\right).$$

The case $\mu \gg N^{-2/m}$ agrees with Armitage-Doll (1954), but $P(\tau_m \le t)$ can grow like Ct^k for any $k = 1, 2, \ldots, m$.

Partial results for general mutation rates

Suppose an individual with $j-1$ mutations acquires a jth mutation at rate u_j.

Theorem: Suppose Z_1 has the Exponential(1) distribution and

- $Nu_1 \to 0$.
- $Nu_2^{1/2}u_3^{1/4}\dots u_m^{1/2^{m-1}} \to \infty$.
- For $j = 1, \dots, m-1$, there is a b_j such that $u_j/u_{j-1} > b_j$ for all N.
- There is an $a > 0$ such that $N^a u_m \to 0$.

Then $Nu_1 u_2^{1/2}u_3^{1/4}\dots u_m^{1/2^{m-1}}\tau_m \to_d Z_1$.

We have a result when $Nu_1 \to 0$ and $Nu_2^{1/2}u_3^{1/4}\dots u_m^{1/2^{m-1}} \to C$. In this case fixation may or may not occur.

We do not have results for general u_j when $Nu_1 \to \infty$.

Open Problems

Complete results for general mutation rates u_j.

Allow individuals with mutations to have a selective advantage.

- When μ is large, model as supercritical, multitype branching process (Moolgavkar-Dewanji-Venzon (1988), Moolgavkar-Luebeck (1990, 1992)).
- When μ is small, fixations are possible. Assume individual with j mutations selected with probability proportional to $1+js$. Beerenwinkel et. al. (2007) conjecture traveling wave behavior. If $Y_j(t)$ denotes the fraction of the population with j mutations at time t, then

$$Y_j(t) \approx C \exp\left(- \frac{(j - vt)^2}{2\sigma^2} \right)$$

and

$$v \approx \frac{2s \log N}{\log(s/\mu)^2}.$$