

II. Exponentially Growing Cell Populations

1 Models

We consider an exponentially growing cell populations modeled as a multi-type branching process in which type i cells are those with $i \geq 0$ mutations and we let $Z_i(t)$ be the number of type i cells at time t . Type i cells give birth at rate a_i and die at rate b_i , where the growth rate $\lambda_i = a_i - b_i > 0$. Thinking of cancer we will usually restrict our attention to the case in which $i \rightarrow \lambda_i$ is increasing. To take care of mutations we suppose that individuals of type i in addition give birth at rate u_{i+1} to individuals of type $i + 1$. This is slightly different than the approach of having mutations with probability u_{i+1} at birth, which translates into a mutation rate of $a_i u_{i+1}$, and this must be kept in mind when comparing results.

The continuous time formulation seems natural for a population of cells with asynchronous reproduction. However, in order to make connection with other results that have been published, we will also consider the discrete time branching processes, which have the advantage that they are easier to simulate. In the model of Bozic et al. (2010), at each time step a cell of type $j \geq 1$ either divides into two cells which occurs with probability b_j , or dies with probability d_j where $d_j = (1 - s)^j/2$ and $b_j = 1 - d_j$. In addition at every division, the new daughter cells can acquire an additional mutation with probability u . In comparing with continuous time, we have to remember that in discrete time the initial exponentially growing population consist of 1's not 0's. This change of notation is somewhat annoying but it allows us to more easily compare with the results in Bozic et al. (2010).

To compute the transition probability of this Markov chain, let $N_j(t)$ be the number of cells of type j . If we let B_j be the number of births, D_j be the number of deaths and M_j be the number of mutations from cells of type j then their joint distribution is multinomial:

$$P[(B_j, D_j, M_j) = (n_1, n_2, n_3)] = \frac{N_j(t)!}{n_1!n_2!n_3!} (b_j(1-u))^{n_1} d_j^{n_2} (b_j u)^{n_3}$$

for $n_1 + n_2 + n_3 = N_j(t)$ and 0 otherwise. Given these variables the size of the next generation is:

$$N_j(t+1) = N_j(t) + B_j - D_j + M_{j-1}$$

2 Branching process results

Continuous time. The number of type-0 cells, $Z_0(t)$, is a branching process. Since each initial individual gives rise to an independent copy of the branching process we will suppose throughout this section that $Z_0(0) = 1$. We begin by computing the extinction probability,

$$\rho = P(Z_0(t) = 0 \text{ for some } t \geq 0).$$

By considering what happened on the first jump

$$\rho = \frac{b_0}{a_0 + b_0} \cdot 1 + \frac{a_0}{a_0 + b_0} \cdot \rho^2$$

Rearranging gives $a_0\rho^2 - (a_0 + b_0)\rho + b_0 = 0$. Since 1 is a root, the quadratic factors as $(\rho - 1)(a_0\rho - b_0) = 0$, and

$$\rho = \begin{cases} b_0/a_0 & \text{if } a_0 > b_0 \\ 1 & \text{if } a_0 \leq b_0 \end{cases} \quad (1)$$

To compute the generating function $F(x, t) = Ex^{Z_0(t)}$, we begin by noting that

Lemma 1. $\partial F/\partial t = -(a_0 + b_0)F + a_0F^2 + b_0$.

Proof. If h is small then the probability of more than one event in $[0, h]$ is $O(h^2)$, the probability of a birth is $\sim a_0h$, of a death is $\sim b_0h$. In the second case we have no particles so the generating function of $Z_0(t + h)$ will be $\equiv 1$. In the first case we have two particles at time h who give rise to two independent copies of the branching process so the generating function of $Z_0(t + h)$ will be $F(x, t)^2$. Combining these observations:

$$F(x, t + h) = a_0hF(x, t)^2 + b_0h \cdot 1 + (1 - (a_0 + b_0)h)F(x, t) + O(h^2)$$

A little algebra converts this into

$$\frac{F(x, t + h) - F(x, t)}{h} = a_0F(x, t)^2 + b_0 - (a_0 + b_0)F(x, t) + O(h)$$

Letting $h \rightarrow 0$ gives the desired result. □

On page 109 of Athreya and Ney (1972), or in formula (5) of Iwasa, Nowak, and Michor (2006) we find the solution:

$$F(x, t) = \frac{b_0(x - 1) - e^{-\lambda_0 t}(a_0x - b_0)}{a_0(x - 1) - e^{-\lambda_0 t}(a_0x - b_0)} \quad (2)$$

which readers skilled at calculus and having sufficient patience can check. Subtracting this from 1 gives

$$1 - F(x, t) = \frac{\lambda_0(x - 1)}{a_0(x - 1) - e^{-\lambda_0 t}(a_0 x - b_0)} \quad (3)$$

Setting $x = 0$, we have

$$\begin{aligned} P(Z_0(t) = 0) &= \frac{b_0 - b_0 e^{-\lambda_0 t}}{a_0 - b_0 e^{-\lambda_0 t}} \\ P(Z_0(t) > 0) &= 1 - F(0, t) = \frac{\lambda_0}{a_0 - b_0 e^{-\lambda_0 t}} \end{aligned} \quad (4)$$

Theorem 1. *Suppose $a_0 > b_0$. As $t \rightarrow \infty$, $e^{-\lambda_0 t} Z_0(t) \rightarrow W_0$ which has Laplace transform*

$$Ee^{-\theta W_0} = \frac{b_0}{a_0} + \left(1 - \frac{b_0}{a_0}\right) \frac{1 - b_0/a_0}{1 - b_0/a_0 + \theta} \quad (5)$$

Hence if δ_0 is a pointmass at 0,

$$W_0 =_d \frac{b_0}{a_0} \delta_0 + \frac{\lambda_0}{a_0} \text{exponential}(\lambda_0/a_0) \quad (6)$$

where the exponential(r) distribution has density re^{-rt} and mean $1/r$.

Proof. The mean $\mu(t) = EZ_0(t)$ satisfies $\mu'(t) = \lambda_0 \mu(t)$ so $EZ_0(t) = e^{\lambda_0 t}$. From this we see that $e^{-\lambda_0 t} Z_0(t)$ is a nonnegative martingale and hence converges to a limit W_0 . To compute the Laplace transform $Ee^{-\theta W_0}$, we set $x = \exp(-\theta e^{-\lambda_0 t})$ in (??) to get

$$\frac{b_0(\exp(-\theta e^{-\lambda_0 t}) - 1) - e^{-\lambda_0 t}(a_0 \exp(-\theta e^{-\lambda_0 t}) - b_0)}{a_0(\exp(-\theta e^{-\lambda_0 t}) - 1) - e^{-\lambda_0 t}(a_0 \exp(-\theta e^{-\lambda_0 t}) - b_0)}$$

As $t \rightarrow \infty$, $e^{-\lambda_0 t} \rightarrow 0$, so $\exp(-\theta e^{-\lambda_0 t}) \rightarrow 1$, $\exp(-\theta e^{-\lambda_0 t}) - 1 \sim -\theta e^{-\lambda_0 t}$, and the above simplifies to

$$\approx \frac{-b_0 \theta e^{-\lambda_0 t} - e^{-\lambda_0 t} \lambda_0}{-a_0 \theta e^{-\lambda_0 t} - e^{-\lambda_0 t} \lambda_0} = \frac{b_0 \theta + \lambda_0}{a_0 \theta + \lambda_0}$$

Dividing top and bottom of this by a_0 and recalling $\lambda_0 = a_0 - b_0$ we have

$$= \frac{(b_0/a_0)\theta + 1 - (b_0/a_0)}{\theta + 1 - (b_0/a_0)} = \frac{b_0}{a_0} + \left(1 - \frac{b_0}{a_0}\right) \frac{1 - (b_0/a_0)}{\theta + 1 - (b_0/a_0)}$$

To invert the Laplace transform, we note that if δ_0 is the point mass at 0 then $p\delta_0 + (1 - p)\text{exponential}(\nu)$ has Laplace transform

$$p + (1 - p) \frac{\nu}{\nu + \theta} = \frac{p\theta + \nu}{\theta + \nu}$$

so $p = b_0/a_0$ and $\nu = 1 - (b_0/a_0)$. □

If we let $\Omega_0^0 = \{Z_0(t) = 0 \text{ for some } t \geq 0\}$ then (??) implies

$$P(\Omega_0^0) = b_0/a_0 \quad (7)$$

Since $W_0 = 0$ on Ω_0^0 , (??) implies that $W_0 > 0$ when the process does not die out. Letting $\Omega_\infty^0 = \{Z_0(t) > 0 \text{ for all } t \geq 0\}$ we have

$$(e^{-\lambda_0 t} Z_0(t) | \Omega_\infty^0) \rightarrow V_0 = \text{exponential}(\lambda_0/a_0) \quad (8)$$

and hence the Laplace transform

$$Ee^{-\theta V_0} = \frac{\lambda_0}{\lambda_0 + a_0 \theta} = (1 + (a_0/\lambda_0)\theta)^{-1}. \quad (9)$$

Discrete time. In this case the 1's are a Galton-Watson process with offspring distribution $p_0 = d_1$, $p_1 = b_1 u$ and $p_2 = b_1(1-u)$. By thinking about what happens on the first step when we start from one cell, one can see that the extinction probability q_1 satisfies

$$q_1 = d_1 + b_1 u q_1 + b_1(1-u)q_1^2$$

Since b_1 and d_1 are $\approx 1/2$ while u is small, e.g., 10^{-4} we have

$$q_1 \approx d_1 + b_1 q_1^2$$

$q_1 = 1$ is a root so factoring $(b_1 \rho - d_1)(\rho - 1)$ we see that

$$q_1 \approx d_1/b_1 \quad (10)$$

Remembering that the new b_1 is the old a_0 and the new d_1 is the old b_0 this is the same as (??). The mean

$$\mu_1 = 2b_1(1-u) + b_1 u \approx 2b_1$$

since u is small. Again $Z_1(n)/\mu^n$ is a martingale and hence $Z_1(n)/\mu^n \rightarrow W_1$. Suppose that $Z_1(0) = 1$. The fact that martingale is L^2 bounded implies $EW_1 = 1$. Thus if we let $V_1 = (W_1 | \Omega_\infty^0)$ then $EV_1 = 1/q_1$.

While the mean is easy to calculate, in discrete time we have very little information about the distribution of W_0 . Imitating the calculation above we see that the Laplace transform $F(\theta, n) = E \exp(-\theta Z_0(n))$ satisfies

$$F(\theta, n) = b_1 F(\theta, n_1)^2 + d_1$$

so if we let $G(\theta) = \lim_{n \rightarrow \infty} E \exp(-\theta Z_1(n)/\mu^n) = \lim_{n \rightarrow \infty} F(\theta/\mu^n, n)$ then

$$G(\theta) = b_1 G(\theta/\mu) + d_1$$

3 Time to reach size M

While from the point of view of stochastic processes it is natural to start measuring time when there is one cancer cell, that time is not known in reality. Thus we will shift our attention to the time at which the cancer is detected, which we will idealize as the time the total number of cancer cells reaches size M . As a first step in investigating this quantity we consider $T_M = \min\{t : Z_0(t) = M\}$, and then return later to consider $Z_i(T_M)$ for $i > 0$.

To find the distribution of T_M , we note that by (??) conditional on nonextinction, $e^{-\lambda_0 t} Z_0(t) \rightarrow V_0$, which is exponential with rate λ_0/a_0 , or informally $Z_0(t) \approx e^{\lambda_0 t} V_0$. From this we see that

$$P(T_M \leq t) = P(e^{\lambda_0 t} V_0 \geq M) = \exp(-(\lambda_0/a_0) M e^{-\lambda_0 t})$$

which is the double exponential, or Gumbel distribution. Differentiating we find the density function

$$f_{T_M}(t) = \exp(-(\lambda_0/a_0) M e^{-\lambda_0 t}) \cdot \frac{\lambda_0^2 M}{a_0} e^{-\lambda_0 t} \quad (11)$$

Clearly $T_M \geq 0$, however $P(T_M \leq 0) = \exp(-\lambda_0 M/a_0)$ which is small in most of our applications so it is natural to view the density in (??) as defined on $(-\infty, \infty)$. To compute the mean we substitute $\exp(-\lambda_0 t) = a_0 z / \lambda_0 M$ or $t = -(1/\lambda_0) \log(a_0 z / \lambda_0 M)$, $dt = -dz/z\lambda_0$ to get

$$\begin{aligned} ET_M &= \frac{\lambda_0^2 M}{a_0} \int_{-\infty}^{\infty} t e^{-\lambda_0 t} \exp(-(\lambda_0/a_0) M e^{-\lambda_0 t}) dt \\ &= -\frac{1}{\lambda_0} \int_0^{\infty} \log(a_0 z / \lambda_0 M) e^{-z} dz \end{aligned}$$

To help see this note that $(\lambda_0^2 M/a_0) e^{-\lambda_0 t} = z\lambda_0$. Since $\int_0^{\infty} e^{-z} dz = 1$ it follows that

$$ET_M = \frac{1}{\lambda_0} \log\left(\frac{M\lambda_0}{a_0}\right) - \frac{1}{\lambda_0} \int_0^{\infty} e^{-z} \log z dz \quad (12)$$

The first term is value of T_M if we replace V_0 by its mean a_0/λ_0 and solve

$$e^{\lambda_0 t} a_0/\lambda_0 = M$$

The second term (including the minus sign) is Euler's constant

$$\gamma = 0.5772156649$$

which is a small correction resulting from the randomness of V_0 .

Example 1. For a concrete example suppose $a_0 = 1.02$, $b = 1$, $\lambda_0 = 0.02$ and set $M = 10^5$. In this case $P(T_M \leq 0) = \exp(-200/1.02) \approx 0$. The first term in (??) is

$$\frac{1}{\lambda_0} \log\left(\frac{M\lambda_0}{a_0}\right) = 50 \log 1960.78 = 379.05$$

so the second is an insignificant correction.

4 Time until the first type 1

Continuous time. Let τ_1 be the time of occurrence of the first type 1. Since 1's are produced at rate $u_1 Z_0(t)$,

$$P(\tau_1 > t | Z_0(s), s \leq t, \Omega_\infty^0) = \exp\left(-u_1 \int_0^t Z_0(s) ds\right) \quad (13)$$

τ_1 will occur when $\int_0^t Z_0(s) ds$ is of order $1/u_1$. A typical choice for $u_1 = 10^{-5}$ or smaller, so $1/u_1$ is a large number, and we can use the approximation $(Z_0(s) | \Omega_\infty^0) \approx e^{\lambda_0 s} V_0$. Evaluating the integral, taking the expected value, and using (??), we conclude that

$$\begin{aligned} P(\tau_1 > t | \Omega_\infty^0) &\approx E \exp(-u_1 V_0 e^{\lambda_0 t} / \lambda_0) \\ &= \frac{\lambda_0}{\lambda_0 + a_0 u_1 e^{\lambda_0 t} / \lambda_0} = (1 + (a_0 / \lambda_0^2) u_1 e^{\lambda_0 t})^{-1} \end{aligned} \quad (14)$$

The median $t_{1/2}^1$ of the distribution has $\lambda_0^2 = a_0 u_1 e^{\lambda_0 t_{1/2}^1}$ so

$$t_{1/2}^1 \approx \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2}{a_0 u_1}\right) \quad (15)$$

In some cases we regard V_0 as a fixed constant. Implicitly assuming that $V_0 > 0$ we write

$$P(\tau_1 > t | V_0) \approx \exp(-u_1 V_0 e^{\lambda_0 t} / \lambda_0).$$

If we replace V_0 by its mean $EV_0 = a_0 / \lambda_0$ the tail of the distribution of τ_1 is equal to $1/e$ at

$$\bar{t}_{1/e}^1 \approx \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2}{a_0 u_1}\right) \quad (16)$$

A second quantity of interest is σ_1 , the time of occurrence of the first type 1 that gives rise to a family which does not die out. Since the rate of these successful type 1 mutations is $u_1 \lambda_1 / a_1$, all we have to do is to replace u_1 by $u_1 \lambda_1 / a_1$ in either (??) or (??), so replacing t by s to define the corresponding quantities for σ_1

$$s_{1/2}^1 = \bar{s}_{1/e}^1 = \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2 a_1}{a_0 u_1 \lambda_1}\right) \quad (17)$$

Example 2. To help digest these formulas it is useful to have concrete examples. If the mutation rate $u_1 = 10^{-5}$, $b_0 = b_1 = 1$, $a_0 = 1.02$, and $a_1 = 1.04$ then $\lambda_0 = 0.02$, $\lambda_1 = 0.04$ and

$$\begin{aligned} t_{1/2}^1 = \bar{t}_{1/2}^1 &= 50 \log\left(\frac{4 \times 10^{-4}}{1.02 \times 10^{-5}}\right) = 183.45 \\ s_{1/2}^1 = \bar{s}_{1/e}^1 &= 50 \log\left(\frac{4.16 \times 10^{-4}}{4.08 \times 10^{-7}}\right) = 50 \log(1019.6) = 346.36 \end{aligned}$$

Discrete time. Specializing the computations in Section 2 of the appendix of Bozic et al (2010) to $j = 1$ and noting that the mean number of offspring of type 1 is $b_1(2 - u)$, since only the new daughter cell is subject to mutation with probability u , we see that the average number of type 1 cells grows as

$$x(n) = \frac{[b_1(2 - u)]^n}{1 - q_1}$$

where n is the number of generations. To convert the number of generations into clock time, we need to take account of the average time per generation, which they call T , but for simplicity, we will omit the conversion from the formulas. New type 2 cells with surviving lineages appear at time $n + 1$ with probability

$$x(n)b_1u(1 - q_2)$$

where q_2 is the extinction probability computed in (??) with the subscript 1 replaced by 2. We approximate the generation number σ_2 of appearance of the first type 2 cell with a surviving lineage by the time when the sum of the probabilities reaches 1, that is, when

$$\sum_{m=0}^{\sigma_2} \frac{[b_1(2 - u)]^m}{1 - q_1} b_1u(1 - q_2) = 1$$

We are assuming $b_1(2 - u) > 1$, so looking backwards from time σ_2 the sum is

$$\approx \frac{[b_1(2 - u)]^{\sigma_2}}{1 - 1/b_1(2 - u)} \cdot \frac{b_1u(1 - q_2)}{1 - q_1} \quad (18)$$

A little algebra gives

$$[b_1(2 - u)]^{\sigma_2} \approx \frac{1 - q_1}{b_1u(1 - q_2)} \cdot \left(1 - \frac{1}{b_1(2 - u)}\right)$$

which, since $\sigma_1 = 0$, leads to

$$\sigma_2 - \sigma_1 \approx \frac{\log\left(\frac{1 - q_1}{b_1u(1 - q_2)} \cdot \left(1 - \frac{1}{b_1(2 - u)}\right)\right)}{\log(b_1(2 - u))} \quad (19)$$

Reintroducing the mean time per generation T , we have a result that is almost identical to (S5) in Bozic et al. (2010) with $j = 1$:

$$\sigma_2 - \sigma_1 \approx \frac{T \log\left(1 + \frac{1 - q_1}{b_1u(1 - q_2)} \cdot \left(1 - \frac{1}{b_1(2 - u)}\right)\right)}{\log(b_1(2 - u))}$$

It is not clear where the 1 comes from but it does not do much to the right-hand side, and it disappears from their subsequent calculations.

To connect with the continuous time formula in (??) we note that the growth rate of the initial population of cells $\lambda_0 = b_1(2 - u) - 1$, so ignoring the $-u$,

$$b_1 = \frac{1 + \lambda_0}{2}, \quad d_1 = 1 - b_1 = \frac{1 - \lambda_0}{2}, \quad \text{and} \quad 1 - q_1 = \frac{b_1 - d_1}{b_1} = \frac{\lambda_0}{b_1}$$

To relate (??) to (??) it is useful to first revisit (??). Substituting the new notation and letting $u_1 = b_1 u$, since mutation only occurs at birth, the sum becomes

$$\approx \frac{(1 + \lambda_0)^{\sigma_2}}{\lambda_0} \cdot \frac{u_1(1 - q_2)}{\lambda_0/b_1}$$

Setting the last quantity equal to 1, and recalling $\sigma_1 = 0$, leads to

$$\sigma_2 - \sigma_1 \approx \frac{\log\left(\frac{\lambda_0^2}{b_1 u_1 (1 - q_1)}\right)}{\log(1 + \lambda_0)}$$

Using $\log(1 + \lambda_0) \approx \lambda_0$ and $1 - q_2 = \lambda_1/b_2$ gives

$$\sigma_2 - \sigma_1 \approx \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2 b_2}{b_1 u_1 \lambda_1}\right) \quad (20)$$

which after converting the b_i 's to a_{i-1} 's is (??).

Example 3. To have a discrete time example similar to Example ?? suppose $b_1 = 0.51$, $d_1 = 0.49$, $b_2 = 0.52$, $d_2 = 0.48$, and $u = 2 \times 10^{-5}$ (since the mutation rate is $u_1 = b_1 u$). In this case $q_1 = 0.49/0.51$ and $q_2 = 0.48/0.52$. Using $1 - 1/b_1(2 - u) \approx 1 - 1/1.02 = 0.0196$, and $1/\log(1.02) = 50.49$ then (??) we have

$$\frac{1 - q_1}{b_1 u (1 - q_2)} \cdot \left(1 - \frac{1}{b_1(2 - u)}\right) = \frac{(0.02)(0.52)}{(0.51)^2(2 \times 10^{-5})(0.04)} \cdot 0.0196 = 980.02$$

$$\sigma_2 - \sigma_1 \approx 50.59 \log(980.02) = 347.75$$

which is almost the same as the answer $50 \log(1019.6) = 364.35$ in Example ?. The difference arises from approximations such as $\log(1 + \lambda_0) \approx \lambda_0$ and $1 - 1/(1 + \lambda_0) \approx \lambda_0$. The factor of two difference in the overall birth rate does not matter because we have adjusted the mutation rate.

Limit Theorems. Our next goal is to find the limiting behavior of τ_1 . For simplicity we consider only continuous time. Since the median is where the distribution function crosses 1/2, (??) implies

$$P(\tau_1 > t_{1/2}^1 + t | \Omega_\infty^0) \approx (1 + e^{\lambda_0 t})^{-1}$$

and it follows that

$$P(\tau_1 > t_{1/2}^1 + x/\lambda_0 | \Omega_\infty^0) \rightarrow (1 + e^x)^{-1} \quad (21)$$

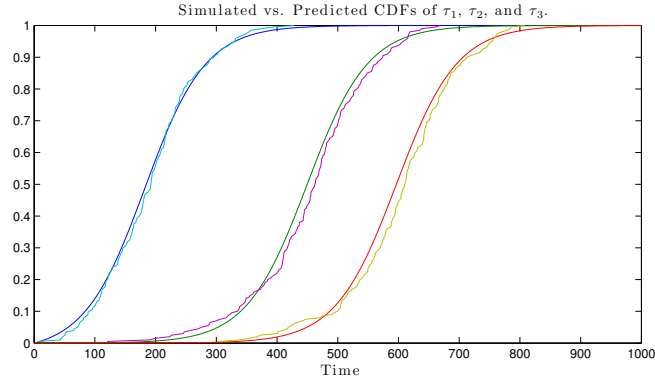


Figure 1: Results of 200 runs of the system with $a_0 = 1.02$, $a_1 = 1.04$, $a_2 = 1.06$, $b_i = 1.0$, and $u = 10^{-5}$. Smooth curves are the limit results for τ_i , $i = 1, 2, 3$.

For a comparison with simulation see Figure ???. The results for fixed V_0 are similar but the limit distributions is slightly different.

$$P(\tau_1 > \bar{t}_{1/e}^1 + t|V_0) \approx \exp(-e^{\lambda_0 t})$$

and it follows that

$$P(\tau_1 > t_{1/2}^1 + x/\lambda_0|V_0) \rightarrow \exp(-e^x)$$

The results for σ_1 come from changing the value of $u_1 \rightarrow u_1 \lambda_2 / a_2$.

5 Mutation before detection?

Iwasa, Nowak, and Michor (2006) were interested in the probability that a mutation conferring resistance to a particular treatment would occur before a cancer was detected. To formulate this a math problem, let $T_M = \min\{t : Z_0(t) = M\}$. Using the calculation in (??), and noting that on the nonextinction event Ω_∞^0 , we have $Z_0(t) \sim V_0 e^{\lambda_0 t}$ implies $Z_0(T_M - s) \approx M e^{-\lambda_0 s}$, we find

$$\begin{aligned} P(\tau_1 > T_M | Z_0(s), s \leq T_M, \Omega_\infty^0) &= \exp\left(-u_1 \int_0^{T_M} Z_0(t) dt\right) \\ &\approx \exp\left(-M u_1 \int_0^\infty e^{-\lambda_0 s} ds\right) = \exp(-M u_1 / \lambda_0) \end{aligned} \quad (22)$$

This answers our math question, but since the mutation to type 1 might die out, the biologically relevant question is to compute the probability that $Z_1(T_M) > 0$. To do this we note that mutations to type-1 occur at rate $u_1 M e^{-\lambda_0 s}$ and by (??) will not die out by time T_M with probability $\lambda_1 / (a_1 - b_1 e^{-\lambda_1 s})$. The number of mutations to type-1 that survive to time T_M is Poisson with mean

$$\mu(M) = M u_1 \int_0^\infty e^{-\lambda_0 s} \frac{\lambda_1}{a_1 - b_1 e^{-\lambda_1 s}} ds \quad (23)$$

and it follows that

$$P(Z_1(T_M) = 0 | \Omega_\infty^0) = \exp(-\mu(M))$$

The integral in (??) cannot be evaluated exactly but it is useful to change variables $t = \exp(-\lambda_0 s)$, $dt = -\lambda_0 \exp(-\lambda_0 s) ds$ to rewrite it as

$$\mu(M) = \frac{M u_1}{\lambda_0} \int_0^1 \frac{\lambda_1}{a_1 - b_1 t^{\lambda_1 / \lambda_0}} dt \quad (24)$$

Logic tells us that $P(\tau_1 > T_M | \Omega_\infty^0) \leq P(Z_1(T_M) = 0 | \Omega_\infty^0)$, so it is comforting to note that the integrand in (??) is ≤ 1 .

To get an upper bound we note that $Z_1(T_M) = 0$ implies $\sigma_1 > T_M$ and type 1 mutations live forever with probability λ_1 / a_1 , so using the reasoning that led to (??)

$$P(Z_1(T_M) = 0 | \Omega_\infty^0) \leq P(\sigma_1 > T_M | \Omega_\infty^0) = \exp(-M u_1 \lambda_1 / a_1 \lambda_0) \quad (25)$$

an inequality which can also be derived by noting that λ_1 / a_1 is a lower bound on the integrand.

Example 4. Leder et al. (2011) compute the probability of pre-existing resistance in chronic myeloid leukemia. They choose $M = 10^5$ cells as the threshold for detection, and on the basis of in vitro studies set $a_0 = 0.008$, $b_0 = 0.003$, and $\lambda = 0.005$ with time measured in years. They are interested in particular nucleotide substitutions, so

they set the mutation rate per birth at 10^{-7} , or at rate $u = 0.008 \cdot 10^{-7} = 8 \times 10^{-10}$. In the case of the mutation *T315I*, which changes Threonine to Isoleucine at position 315, the growth parameters are $a_1 = 0.0088$ and $b_1 = 0.003$, so $\lambda_1 = 0.0058$. In this case $Mu_1/\lambda_0 = 10^5 \cdot 0.008 \cdot 10^{-7}/0.005 = 0.016$, and $\lambda_1/a_1 = 0.659$, so we have

$$\begin{aligned} P(\tau_1 \leq T_M | \Omega_\infty^0) &= 1 - e^{-0.016} = 0.015873 \\ P(\sigma_1 \leq T_M | \Omega_\infty) &= 1 - e^{-0.010544} = 0.01049 \\ P(Z_1(T_M) > 0 | \Omega_\infty) &= 0.01263 \end{aligned}$$

where the last answer comes from evaluating the integral in (??) numerically. The mutation with the lowest growth rate that they consider is L248R, which changes Leucine to Arginine at position 248. It has growth parameters $a_1 = 0.0061$ and $b_1 = 0.003$, so $\lambda_1 = 0.0031$, $\lambda_1/a_1 = 0.581$ and the results in this case are

$$\begin{aligned} P(\tau_1 \leq T_M | \Omega_\infty^0) &= 1 - e^{-0.016} = 0.015873 \\ P(\sigma_1 \leq T_M | \Omega_\infty) &= 1 - e^{-0.008131} = 0.00810 \\ P(Z_1(T_M) > 0 | \Omega_\infty) &= 0.01198 \end{aligned}$$

6 Properties of the Gamma Function

The Gamma function is defined for $\alpha > 0$ by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (26)$$

This quantity with $0 < \alpha < 1$ will show up in the constants of our limit theorems, so we record some of its properties now. Integrating by parts

$$\Gamma(\alpha + 1) = \int_0^{\infty} t^{\alpha} e^{-t} dt = \int_0^{\infty} \alpha t^{\alpha-1} e^{-t} dt = \alpha \Gamma(\alpha) \quad (27)$$

Since $\Gamma(1) = 1$ it follows that if n is an integer $\Gamma(n) = (n - 1)!$. Among the many formulas for Γ , the most useful for us is *Euler's reflection formula*

$$\Gamma(\alpha)\Gamma(1 - \alpha) = \frac{\pi}{\sin(\pi\alpha)} \quad (28)$$

Taking $\alpha = 1/2$ we see that implies $\Gamma(1/2) = \sqrt{\pi}$. Letting $\alpha \rightarrow 0$ and using $\Gamma(1-\alpha) \rightarrow 1$

$$\Gamma(\alpha) \sim \frac{\pi}{\sin(\pi\alpha)} \sim \frac{1}{\alpha} \quad (29)$$

where we have used $\sin x \sim x$ as $x \rightarrow 0$.

7 Growth of $Z_1(t)$

In this section we will examine the growth of the type 1's under the assumption that $Z_0^*(t) = V_0 e^{\lambda_0 t}$ for $t \in (-\infty, \infty)$. Here the star is to remind us that we have extended Z_0 to negative times.

Theorem 2. *If we assume $Z_0^*(t) = V_0 e^{\lambda_0 t}$ then as $t \rightarrow \infty$, $e^{-\lambda_1 t} Z_1^*(t) \rightarrow V_1$ where V_1 is the sum of the points in a Poisson process with mean measure $\mu(x, \infty) = c_{\mu,1} u_1 V_0 x^{-\alpha}$ where $\alpha = \lambda_0/\lambda_1$,*

$$c_{\mu,1} = \frac{1}{a_1} \left(\frac{a_1}{\lambda_1} \right)^\alpha \Gamma(\alpha) \quad (30)$$

Here and what follows constants like $c_{\mu,1}$ will depend on the branching process parameters a_i and b_i , but not on the mutation rates u_i .

Proof. Mutations to type 1 occur at times of a Poisson process with rate $u_1 V_0 e^{\lambda_0 s}$. Theorem ?? implies a mutation at time s will grow to size $\approx e^{\lambda_1(t-s)} W_1$ by time t , where W_1 has distribution

$$W_1 =_d \frac{b_1}{a_1} \delta_0 + \frac{\lambda_1}{a_1} \text{exponential}(\lambda_1/a_1)$$

To add up the contributions, we associate with each point s_i in the Poisson process an independent random variable y_i with the same distribution as W_1 . This gives us a Poisson process which on $(-\infty, \infty) \times (0, \infty)$ (we ignore the points with $y_i = 0$) has intensity

$$u_1 V_0 e^{\lambda_0 s} \cdot (\lambda_1/a_1)^2 e^{-(\lambda_1/a_1)y}$$

Here, one the two factors of λ_1/a_1 comes from $P(W_1 > 0)$, the other from the exponential density function.

A point (s, y) makes a contribution $e^{-\lambda_1 s} y$ to $\lim_{t \rightarrow \infty} e^{-\lambda_1 t} Z_1^*(t)$. Points with $e^{-\lambda_1 s} y > x$ will contribute more than x to the limit. The number of such points is Poisson distributed with mean

$$\int_{-\infty}^{\infty} u_1 V_0 e^{\lambda_0 s} \frac{\lambda_1}{a_1} e^{-(\lambda_1/a_1)x e^{\lambda_1 s}} ds$$

where one factor of λ_1/a_1 has disappeared since we are looking at the tail of the distribution. Changing variables

$$\frac{\lambda_1}{a_1} x e^{\lambda_1 s} = t, \quad \frac{\lambda_1}{a_1} x \lambda_1 e^{\lambda_1 s} ds = dt$$

and noticing $s = (1/\lambda_1) \log(t a_1/x \lambda_1)$ implies $e^{(\lambda_0 - \lambda_1)s} = (a_1 t/\lambda_1 x)^{(\lambda_0/\lambda_1) - 1}$ the integral above becomes

$$\begin{aligned} &= u_1 V_0 \int_0^\infty \left(\frac{a_1 t}{\lambda_1 x} \right)^{(\lambda_0/\lambda_1) - 1} e^{-t} \frac{dt}{\lambda_1 x} \\ &= \frac{u_1 V_0}{a_1} \left(\frac{a_1}{\lambda_1} \right)^{\lambda_0/\lambda_1} x^{-\lambda_0/\lambda_1} \int_0^\infty t^{(\lambda_0/\lambda_1) - 1} e^{-t} dt \end{aligned}$$

which completes the proof. \square

7.1 Laplace transform proof of Theorem ??.

The previous proof is useful for the insights it gives into the limit V_1 , but to analyze the size of later waves using mathematical induction, it is convenient to approach things using Laplace transforms.

Theorem 3. *If we assume $Z_0^*(t) = V_0 e^{\lambda_0 t}$ for $t \in (-\infty, \infty)$ then $e^{-\lambda_1 t} Z_1^*(t) \rightarrow V_1$ as $t \rightarrow \infty$ where V_1 has $E(e^{-\theta V_1} | V_0) = \exp(-c_{h,1} u_1 V_0 \theta^\alpha)$ with*

$$c_{h,1} = c_{\mu,1} \Gamma(1 - \alpha) = \frac{1}{a_1} \left(\frac{a_1}{\lambda_1} \right)^\alpha \Gamma(\alpha) \Gamma(1 - \alpha) \quad (31)$$

If V_0 is exponential(λ_0/a_0) then $E \exp(-\theta V_1) = (1 + c_{h,1} u_1 (a_0/\lambda_0) \theta^\alpha)^{-1}$.

Using the identity $\Gamma(1 + \alpha) = \alpha \Gamma(\alpha)$ from (??) one can see that this is the same as the constant $c_{h,1}$ defined at the bottom of page 43 in Durrett and Moseley (2010).

Proof. The second conclusion follows from the first and (??). Let $\tilde{Z}_1(t)$ be the number of 1's at time t in the branching process with $Z_0(0) = 0$, $Z_1(0) = 1$, and let $\tilde{\phi}_{1,t}(\theta) = E e^{-\theta \tilde{Z}_1(t)}$.

Lemma 2. $E(e^{-\theta Z_1^*(t)} | V_0) = \exp\left(-u_1 \int_{-\infty}^t V_0 e^{\lambda_0 s} (1 - \tilde{\phi}_{1,t-s}(\theta)) ds\right)$

Proof. We begin with the corresponding formula in discrete time:

$$\begin{aligned} E(e^{-\theta Z_1^*(n)} | Z_0(m), m \leq n) &= \prod_{m=-\infty}^{n-1} \sum_{k_m=0}^{\infty} e^{-u_1 Z_0(m)} \frac{(u_1 Z_0(m))^{k_m}}{k_m!} \tilde{\phi}_{1,n-m-1}(\theta)^{k_m} \\ &= \prod_{m=-\infty}^{n-1} \exp\left(-u_1 Z_0(m) (1 - \tilde{\phi}_{1,n-m-1}(\theta))\right) \\ &= \exp\left(-u_1 \sum_{m=-\infty}^{n-1} Z_0(m) (1 - \tilde{\phi}_{1,n-m-1}(\theta))\right) \end{aligned}$$

Breaking up the time-axis into intervals of length h and letting $h \rightarrow 0$ and using $Z_0^*(s) = V_0 e^{\lambda_0 s}$ gives the result in continuous time. \square

Replacing θ by $\theta e^{-\lambda_1 t}$ in Lemma ?? and letting $t \rightarrow \infty$

$$E(e^{-\theta V_1} | V_0) = \lim_{t \rightarrow \infty} \exp\left(-u_1 V_0 \int_{-\infty}^t e^{\lambda_0 s} (1 - \tilde{\phi}_{1,t-s}(\theta e^{-\lambda_1 t})) ds\right) \quad (32)$$

To calculate the limit, we note that by (??)

$$\tilde{Z}_1(t-s)e^{-\lambda_1(t-s)} \Rightarrow \frac{b_1}{a_1}\delta_0 + \frac{\lambda_1}{a_1} \text{exponential}(\lambda_1/a_1)$$

so multiplying by $e^{-\lambda_1 s}$ and taking the Laplace transform, we have

$$1 - \tilde{\phi}_{t-s}(\theta e^{-\lambda_1 t}) \rightarrow \frac{\lambda_1}{a_1} \int_0^\infty (1 - e^{-\theta x})(\lambda_1/a_1)e^{\lambda_1 s} e^{-xe^{\lambda_1 s}\lambda_1/a_1} dx \quad (33)$$

Using this in (??) and interchanging the order of integration

$$E(e^{-\theta V_1} | V_0) = \exp(-u_1 V_0 h(\theta)) \quad (34)$$

where

$$h(\theta) = (\lambda_1^2/a_1^2) \int_0^\infty (1 - e^{-\theta x}) \left[\int_{-\infty}^\infty e^{\lambda_0 s} e^{\lambda_1 s} e^{-xe^{\lambda_1 s}\lambda_1/a_1} ds \right] dx. \quad (35)$$

Changing variables $u = xe^{\lambda_1 s}\lambda_1/a_1$, $e^{\lambda_1 s} ds = a_1 du/(\lambda_1^2 x)$, the inside integral

$$= \int_0^\infty \frac{a_1}{x\lambda_1^2} \left(\frac{a_1 u}{\lambda_1 x} \right)^{\lambda_0/\lambda_1} e^{-u} du$$

Inserting this in (??) and recalling $\alpha = \lambda_0/\lambda_1$, we have

$$h(\theta) = \frac{1}{a_1} \left(\frac{a_1}{\lambda_1} \right)^\alpha \int_0^\infty (1 - e^{-\theta x}) x^{-\alpha-1} dx \int_0^\infty u^\alpha e^{-u} du$$

Comparing with (??) and remembering $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ (integrate by parts) gives

$$h(\theta) = c_{\mu,1} \int_0^\infty (1 - e^{-\theta x}) \alpha x^{-\alpha-1} dx \quad (36)$$

Changing variables $x = y/\theta$, $dx = dy/\theta$ we have

$$h(\theta) = c_{\mu,1} \theta^\alpha \int_0^\infty (1 - e^{-y}) \alpha y^{-\alpha-1} dy$$

Integrating by parts it follows that

$$h(\theta) = c_{\mu,1} \theta^\alpha \int_0^\infty e^{-y} y^{-\alpha} dy = c_{\mu,1} \Gamma(1-\alpha) \theta^\alpha = c_{h,1} \theta^\alpha \quad (37)$$

which completes the proof of the theorem. \square

Theorems ?? and ?? results show that $e^{-\lambda_1 t} Z_1^*(t) \rightarrow V_1$ so the two descriptions of the limit must coincide. To begin to check this, we note that if S is the sum of Poisson mean λ number of independent random variables with distribution μ then

$$\begin{aligned} Ee^{-\theta S} &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \left(\int e^{-\theta x} \mu(dx) \right)^k \\ &= \exp \left(-\lambda + \lambda \int e^{-\theta x} \mu(dx) \right) \\ &= \exp \left(- \int (1 - e^{-\theta}) \lambda \mu(dx) \right) \end{aligned}$$

Let $B = c_{\mu,1} \alpha u_1 V_0$, $\lambda_\epsilon = \int_\epsilon^\infty Bx^{-\alpha} dx$ and μ_ϵ have density $\lambda_\epsilon^{-1} Bx^{-\alpha}$ on (ϵ, ∞) . If S_ϵ is the sum of Poisson mean λ_ϵ number of independent random variables with distribution μ_ϵ then

$$Ee^{-\theta S_\epsilon} = \exp \left(- \int_\epsilon^\infty (1 - e^{-\theta}) Bx^{-\alpha} dx \right)$$

Letting $\epsilon \rightarrow 0$ we see that if V_1 is the sum of the points in a Poisson process with mean measure $\mu(x, \infty) = Bx^{-\alpha}$ then the V_1 defined in Theorem ?? has

$$Ee^{-\theta V_1} = \exp \left(- \int_0^\infty (1 - e^{-\theta x}) Bx^{-\alpha-1} dx \right)$$

Combining (??) and (??) we see that the two V_1 's have the same distribution.

While the proof of Theorem ?? is fresh in the reader's mind we will prove:

Theorem 4. *As $M \rightarrow \infty$, $Z_1^*(T_M)/(Mu_1)^{\lambda_1/\lambda_0}$ converges to U_1 in distribution where*

$$E(\exp(-\theta U_1)) = \exp(-c_{h,1} u_1 \theta^{\lambda_0/\lambda_1})$$

and $c_{\mu,1}$ is the constant in (??).

Proof. Working backward from T_M , assuming deterministic growth of type-0 cells at rate $e^{\lambda_0 s}$, and using the proof of Lemma ??, we can show

$$E \exp \left(- \frac{\theta Z_1^*(T_M)}{(Mu_1)^{\lambda_1/\lambda_0}} \right) \approx \exp \left(-u_1 \int_{-\infty}^0 M e^{\lambda_0 s} (1 - \tilde{\phi}_{-s}(\theta (Mu_1)^{-\lambda_1/\lambda_0})) ds \right)$$

We are interested in finding

$$\lim_{M \rightarrow \infty} \exp \left[-u_1 \int_{-\infty}^0 M e^{\lambda_0 s} (1 - \tilde{\phi}_{-s}(\theta (Mu_1)^{-\lambda_1/\lambda_0})) ds \right]$$

First, we make the change of variables $s = t - \frac{1}{\lambda_0} \log(Mu_1)$.

$$= \lim_{M \rightarrow \infty} \exp \left[-u_1 \int_{-\infty}^{\frac{1}{\lambda_0} \log(Mu_1)} e^{\lambda_0 t} (1 - \tilde{\phi}_{\frac{1}{\lambda_0} \log(Mu_1) - t}(\theta(Mu_1)^{-\lambda_1/\lambda_0})) dt \right]$$

Taking the limit as $M \rightarrow \infty$ is essentially the same calculation as (??).

$$= \exp \left[-u_1 \int_{-\infty}^{\infty} e^{\lambda_0 t} \frac{\lambda_1}{a_1} \int_0^{\infty} (1 - e^{-\theta x})(\lambda_1/a_1) e^{\lambda_1 t} e^{-x e^{\lambda_1 t} \lambda_1/a_1} dx dt \right]$$

We conclude by recognizing this double integral as $h(\theta)$ defined in (??) and computed in (??). \square

7.2 Power law tail?

To demonstrate the usefulness of having two descriptions of the limit, we will now show

Theorem 5. *Suppose $Z_0(t) = V_0 e^{\lambda_0 t}$ for $t \in (-\infty, \infty)$ then*

$$P(V_1 > x | V_0) \sim c_{\mu,1} u_1 V_0 x^{-\alpha}$$

If V_0 is exponential(λ_0/a_0) then $P(V_1 > x) \sim c_{\mu,1} u_1 (a_0/\lambda_0) x^{-\alpha}$.

Proof. The second result follows from the first by taking expected value. It is clear from Theorem ?? that

$$P(V_1 > x | V_0) \geq c_{\mu,1} u_1 V_0 x^{-\alpha}.$$

To show that V_1 has a power law tail, we note that as $\theta \rightarrow 0$,

$$1 - E(e^{-\theta V_1} | V_0) \sim c_{\mu,1} \Gamma(1 - \alpha) u_1 V_0 \theta^\alpha \quad (38)$$

and then use a Tauberian theorem from Feller Volume II (pages 442–446). Let

$$\omega(\lambda) = \int_0^{\infty} e^{-\lambda x} dU(x)$$

Lemma 3. *If L is slowly varying and U has an ultimately monotone derivative u , then $\omega(\lambda) \sim \lambda^{-\rho} L(1/\lambda)$ if and only if $u(x) \sim x^{\rho-1} L(x)/\Gamma(\rho)$.*

To use this result we note that if $\phi(\theta)$ is the Laplace transform of the probability distribution F , then integrating by parts gives

$$\int_0^{\infty} e^{-\theta x} dF(x) = (e^{-\theta x})(F(x) - 1)|_0^{\infty} - \theta \int_0^{\infty} e^{-\theta x} (1 - F(x)) dx$$

so we have

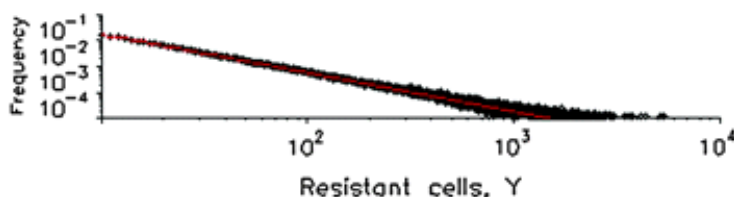
$$1 - \phi(\theta) = \theta \int_0^\infty e^{-\theta x} (1 - F(x)) dx$$

Using (??), it follows that $(1 - \phi(\theta))/\theta \sim c\theta^{\alpha-1}$. Apply Lemma ?? with $\omega(\theta) = (1 - \phi(\theta))/\theta$, $u(x) = 1 - F(x)$ which is decreasing and $\rho = 1 - \alpha$ we conclude

$$1 - F(x) \sim \frac{c}{\Gamma(1 - \alpha)} x^{-\lambda_0/\lambda_1}$$

which proves the desired results. □

The result in Theorem ?? was discovered by Iwasa, Noawk, and Michor (2006) using simulation. The next graph gives part e of their Figure 3.



While the simulation result may be convincing, the power law tail is due to the extension to $(-\infty, 0)$ If we truncate the integral in the proof of Theorem ?? at 0 then the calculation at the end gives

$$\begin{aligned} & \int_0^\infty u_1 V_0 e^{\lambda_0 s} \frac{\lambda_1}{a_1} e^{-(\lambda_1/a_1)x e^{\lambda_1 s}} ds \\ &= \frac{u_1 V_0}{a_1} \left(\frac{a_1}{\lambda_1} \right)^{\lambda_0/\lambda_1} x^{-\lambda_0/\lambda_1} \int_{\lambda_1 x/a_1}^\infty t^{(\lambda_0/\lambda_1)-1} e^{-t} dt \end{aligned}$$

so when $\lambda_1 x/a_1$ is large the tail decays exponentially with a power law correction.

Insert results for mean

8 Growth of $Z_k(t)$ and τ_k for $k \geq 2$

Let $\alpha_k = \lambda_{k-1}/\lambda_k$. Generalizing (??) we define

$$c_{h,k} = \frac{1}{a_k} \left(\frac{a_k}{\lambda_k} \right)^{\alpha_k} \Gamma(\alpha_k) \Gamma(1 - \alpha_k) \quad (39)$$

Let $c_{\theta,0} = a_0/\lambda_0$, $\mu_0 = 1$ and inductively define for $k \geq 1$ $c_{\theta,k} = c_{\theta,k-1} c_{h,k}^{\lambda_0/\lambda_{k-1}}$ and

$$\mu_k = \mu_{k-1} u_k^{\lambda_0/\lambda_{k-1}} = \prod_{j=1}^k u_j^{\lambda_0/\lambda_{j-1}}.$$

Theorem 6. *Suppose $Z_0^*(t) = V_0$ for $t \in (-\infty, \infty)$ where V_0 is exponential(λ_0/a_0). Then $e^{-\lambda_k t} Z_k^*(t) \rightarrow V_k$ a.s. with*

$$E e^{-\theta V_k} = (1 + c_{\theta,k} \mu_k \theta^{\lambda_0/\lambda_k})^{-1} \quad (40)$$

and hence $P(V_k > x) \sim c_{V,k} \mu_k x^{-\lambda_0/\lambda_k}$, where $c_{V,k} = c_{\theta,k}/\Gamma(1 - \lambda_0/\lambda_k)$.

Proof. The result for $P(V_k > x)$ follows from (??) and Lemma ??. We will prove the formula for $E e^{-\theta V_k}$ by induction. When $k = 1$, $c_{\theta,1} = c_{h,1} a_0/\lambda_0$. so this follows from Theorem ??. Suppose now that $k \geq 2$. Let \mathcal{F}_t^{k-1} be the σ -field generated by $Z_j^*(s)$ for $j \leq k-1$ and $s \leq t$. Let $\tilde{Z}_k(t)$ be the number of type k 's at time t in the branching process with $\tilde{Z}_k(0) = 1$ and $\tilde{Z}_j(0) = 0$ for $j \leq k-1$, and let $\tilde{\phi}_{k,t}(\theta) = E e^{-\theta \tilde{Z}_k(t)}$. The reasoning that led to Lemma ?? implies

$$E(e^{-\theta Z_k^*(t)} | \mathcal{F}_t^{k-1}) = \exp \left(-u_k \int_{-\infty}^t Z_{k-1}^*(s) (1 - \tilde{\phi}_{k,t-s}(\theta)) ds \right)$$

Replacing $Z_{k-1}^*(s)$ by $e^{\lambda_{k-1}s} V_{k-1}$, θ by $\theta e^{-\lambda_k t}$, and letting $t \rightarrow \infty$

$$E(e^{-\theta V_k} | \mathcal{F}_\infty^{k-1}) = \lim_{t \rightarrow \infty} \exp \left(-u_k V_{k-1} \int_{-\infty}^t e^{\lambda_{k-1}s} (1 - \tilde{\phi}_{k,t-s}(\theta e^{-\lambda_k t})) ds \right) \quad (41)$$

At this point the calculation is the same as the one in the proof of Theorem ?? with 1 and 0 replaced by k and $k-1$ respectively. Combining (??) and (??) we conclude that

$$E(e^{-\theta V_k} | \mathcal{F}_\infty^{k-1}) = \exp(-u_k V_{k-1} c_{h,k} \theta^{\lambda_{k-1}/\lambda_k}) \quad (42)$$

Taking expected value and using the result for $k-1$

$$\begin{aligned} E e^{-\theta V_k} &= (1 + c_{\theta,k-1} \mu_{k-1} (u_k c_{h,k} \theta^{\lambda_{k-1}/\lambda_k})^{\lambda_0/\lambda_{k-1}})^{-1} \\ &= (1 + c_{\theta,k} \mu_k \theta^{\lambda_0/\lambda_k})^{-1} \end{aligned}$$

which proves the result. □

Our next topic is the waiting time for the first type $k + 1$:

$$P(\tau_{k+1} > t | \mathcal{F}_t^k) = \exp\left(-\int_0^t Z_k^*(s) ds\right) \approx \exp(-u_{k+1} V_k e^{\lambda_k t} / \lambda_k)$$

Taking expected value and using Theorem ??

$$P(\tau_{k+1} > t | \Omega_\infty^0) = (1 + c_{\theta,k} \mu_k (u_{k+1} e^{\lambda_k t} / \lambda_k)^{\lambda_0 / \lambda_k})^{-1}$$

Using the definition of μ_{k+1} the median $t_{1/2}^{k+1}$ is defined by

$$c_{\theta,k} \mu_{k+1} \exp(\lambda_0 t_{1/2}^{k+1}) \lambda_k^{-\lambda_0 / \lambda_k} = 1$$

and solving gives

$$t_{1/2}^{k+1} = \frac{1}{\lambda_0} \log\left(\frac{\lambda_k^{\lambda_0 / \lambda_k}}{c_{\theta,k} \mu_{k+1}}\right) \quad (43)$$

As in the case of τ_1

$$P(\tau_{k+1} > t_{1/2}^{k+1} + x / \lambda_0) \approx (1 + e^x)^{-1}$$

When $k = 0$, $c_{\theta,0} = a_0 / \lambda_0$ and we have the result in (??)

$$t_{1/2}^1 = \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2}{a_0 u_1}\right)$$

To complete the picture, we need to relate $t_{1/2}^{k+1}$ to $t_{1/2}^k$. To do this we use the recursions introduced before Theorem ?? to get

$$\begin{aligned} t_{1/2}^{k+1} &= \frac{1}{\lambda_k} \log \lambda_k + \frac{1}{\lambda_0} \log\left(\frac{1}{c_{\theta,k-1} c_{h,k}^{\lambda_0 / \lambda_{k-1}} \mu_k u_{k+1}^{\lambda_0 / \lambda_k}}\right) \\ &= \frac{1}{\lambda_k} \log\left(\frac{\lambda_k}{u_{k+1}}\right) + \frac{1}{\lambda_{k-1}} \log\left(\frac{1}{c_{h,k} \lambda_{k-1}}\right) + \frac{1}{\lambda_0} \log\left(\frac{\lambda_{k-1}^{\lambda_0 / \lambda_{k-1}}}{c_{\theta,k-1} \mu_k}\right) \end{aligned}$$

Using (??) we have

$$t_{1/2}^{k+1} - t_{1/2}^k = \frac{1}{\lambda_k} \log\left(\frac{\lambda_k^2}{a_k u_{k+1}}\right) + \frac{1}{\lambda_{k-1}} \log\left(\frac{a_k}{\lambda_{k-1} \Gamma(\alpha_k) \Gamma(1 - \alpha_k)}\right) \quad (44)$$

If we are waiting for σ_{k+1} the time of the first successful type $k + 1$ mutation then we need to replace u_{k+1} by $u_{k+1} \lambda_{k+1} / a_{k+1}$ and the recursion for its median becomes

$$\begin{aligned} s_{1/2}^{k+1} - s_{1/2}^k &= \frac{1}{\lambda_k} \log\left(\frac{\lambda_k^2 a_{k+1}}{a_k u_{k+1} \lambda_{k+1}}\right) \\ &\quad + \frac{1}{\lambda_{k-1}} \log\left(\frac{a_k}{\lambda_{k-1} \Gamma(\alpha_k) \Gamma(1 - \alpha_k)}\right) \end{aligned} \quad (45)$$

Discrete time. In contrast (S5) in Bozic et al. (2010) says

$$\sigma_{j+1} - \sigma_j = \frac{\log \left(1 + \frac{1-q_j}{b_j u (1-q_{j+1})} \cdot \left(1 - \frac{1}{b_j (2-u)} \right) \right)}{\log[b_j(2-u)]} \approx \frac{1}{\lambda_j} \log \left(\frac{\lambda_j^2 b_{j+1}}{b_j u_{j+1} \lambda_{j+1}} \right) \quad (46)$$

where in the second step we have used the reasoning that led to (??). The authors do not explain the reasoning that led to this formula, but comparing with (??) the formula implies that that when we shift time to σ_j we do not reduce the problem for waiting for σ_1 . This is not correct. The initial wave is started by one cell while in later waves the first successful type k mutation will soon be followed by others.

We have $d_j = (1-s)^j/2$ and $b_j = 1 - d_j$ so

$$q_j = \frac{d_j}{b_j} = \frac{(1-s)^j/2}{1 - (1-s)^j/2} \approx \frac{1-js}{1+js} \approx 1 - 2js$$

Since $b_j \approx 1/2$, it follows that

$$\frac{1-q_j}{b_j u (1-q_{j+1})} \approx \frac{2j}{u(j+1)}$$

Using $\lambda_j = b_j(2-u) - 1 \approx js$, $\log(b_j(2-u)) \approx \log(1+js) \approx js$ and $1 - 1/(b_j(2-u)) \approx js$, we have (S6) from Bozic et al. (2010)

$$\sigma_{j+1} - \sigma_j \approx \frac{T}{js} \log \left(\frac{2j^2 s}{u(j+1)} \right) \quad (47)$$

where T is the generation time. The apparent extra factor of 2 here compared to (??) is due to the fact that $u_{j+1} = b_j u \approx u/2$. Bozic et al. (2010) suggest that we can simplify this to

$$\sigma_{j+1} - \sigma_j \approx \frac{T}{js} \log \left(\frac{2js}{u} \right) \quad (48)$$

because $\log(j/(j+1))$ is much smaller than $\log(2js/u)$.

Example 5. Bozic et al. (2010) argue that appropriate parameters for colon cancer are $u = 10^{-5}$, $s = 10^{-2}$, and $T = 4$ days. In this case (??) gives

$$\sigma_2 - \sigma_1 = 400 \log(2000) = 3040 \text{ days} = 8.33 \text{ years}$$

$$\sigma_3 - \sigma_2 = 200 \log(4000) = 1659 \text{ days} = 4.54 \text{ years}$$

which agrees with the numbers given on page 18546 of their paper. In contrast (??) gives

$$\sigma_2 - \sigma_1 = 400 \log(1000) = 2763 \text{ days} = 7.57 \text{ years}$$

$$\sigma_3 - \sigma_2 = 200 \log(8000/3) = 1577 \text{ days} = 4.32 \text{ years}$$

so in the first case (??) introduces a significant error. To see what impact that the missing term has in the second case we note that using (??) with $k = 1$, $\lambda_j = (j+1)s$ we have $\alpha_1 = 1/2$, $a_1 = b_1 = 1/2$, and $\Gamma(1/2) = \sqrt{\pi}$ so the missing term is

$$100 \log(50/\pi) = 276.72$$

9 Transitions between waves

In this section we investigate the time $T_k = \inf\{t \geq 0 : Z_k(t) > Z_j(t) \text{ for all } j \neq k\}$ at which the type- k 's first become dominant in the population. Our first step is to remove the mutation rate from the limit distribution.

Theorem 7. *Suppose $Z_0^*(t) = V_0$ for $t \in (-\infty, \infty)$ where V_0 is exponential(λ_0/a_0). Then $\mu_k^{-\lambda_k/\lambda_0} e^{-\lambda_k t} Z_k^*(t) \rightarrow \bar{V}_k$ a.s. with*

$$E e^{-\theta \bar{V}_k} = (1 + c_{\theta,k} \theta^{\lambda_0/\lambda_k})^{-1}$$

Recall $\mu_k = \prod_{j=1}^k u_j^{\lambda_0/\lambda_{j-1}}$. For simplicity we will consider the special case in which all the $u_i = u$ and hence

$$\mu_k^{-\lambda_k/\lambda_0} = (1/u)^{\rho(k)} \quad \text{where} \quad \rho(k) = \sum_{j=0}^{k-1} \frac{\lambda_k}{\lambda_{j-1}}$$

The result in Theorem ?? suggests that for large t

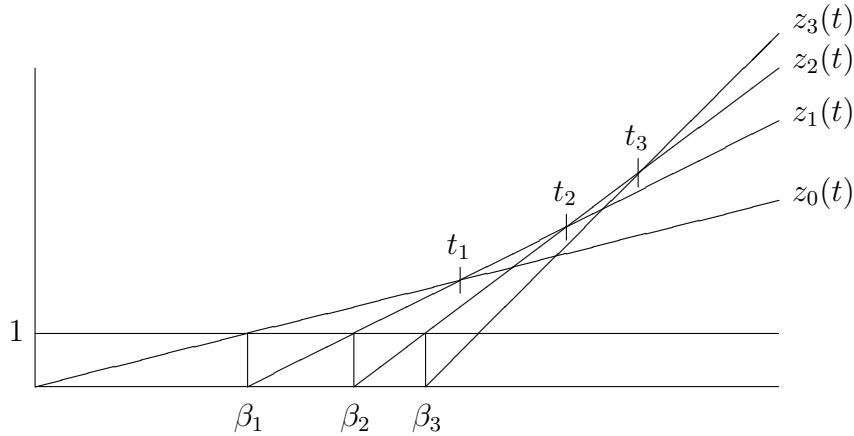
$$\log Z_k(t) \approx \lambda_k t - \rho(k) \log(1/u) + \log(\bar{V}_k)$$

Let $L = \log(1/u)$. If we speed up time by a factor of L , divide both sides by L then we have

Theorem 8. *Let $\beta_k = \rho(k)/\lambda_k = \sum_{j=0}^{k-1} 1/\lambda_j$. As $u \rightarrow 0$*

$$\frac{1}{L} \log^+ Z_k(Lt) \rightarrow z_k(t) = \lambda_k(t - \beta_k)^+$$

Here $x^+ = \max\{0, x\}$ takes care of the fact that $\log(0) = -\infty$. A picture tells the story much better than formulas:



In words $Z_{k-1}(Lt)$ hits $1/u$ at time $\approx \beta_k$. At this point the first type k is born and the population grows like $e^{\lambda_k t}$, i.e., its logarithm grows like $\lambda_k t$. It is clear from the definition that

$$\beta_k - \beta_{k-1} = \frac{1}{\lambda_{k-1}} \quad (49)$$

so the process is accelerating, i.e., the increments between the birth times for successive waves are decreasing.

Wave k overtakes wave $k-1$ at the time $t_k > \beta_k$ when $\lambda_k(t - \beta_k) = \lambda_{k-1}(t - \beta_{k-1})$ or

$$(\lambda_k - \lambda_{k-1})t_k = \lambda_k\beta_k - \lambda_{k-1}\beta_{k-1}$$

In the special case $\lambda_k = \lambda_0 + kb$ this becomes

$$bt_k = b\beta_k + \frac{1}{\lambda_{k-1}}(\beta_k - \beta_{k-1})$$

so using (??)

$$t_k = \beta_k + b^{-1}$$

Note that this is a constant time after the time the first type k appears:

Theorem 9. *If $u_j \equiv u$ and $\lambda_k = \lambda_0 + kb$ then $T_k/L \rightarrow \beta_k + b^{-1}$*

10 Heterogeneity

The point process representation of $V_1 = \lim_{t \rightarrow \infty} e^{-\lambda_1 t} Z_1^*(t)$ in Theorem ?? is very useful for considering tumor heterogeneity. One reason for this is that it allows us to make a connection between V_1 and one-sided stable laws.

Theorem 10. *Let Y_1, Y_2, \dots be independent and identically distributed nonnegative random variables with $P(Y_i > x) \sim cx^{-\alpha}$ where $0 < \alpha < 1$. Let $S_n = Y_1 + \dots + Y_n$. Then*

$$S_n/n^{1/\alpha} \Rightarrow W$$

where W is the sum of the points in a Poisson process with mean measure $\mu(z, \infty) = cz^{-\alpha}$.

Why is this true? $|\{i \leq n : Y_i \geq xn^{-1/\alpha}\}|$ is \approx Binomial($n, cx^{-\alpha}/n$) and hence converges to Poisson($cx^{-\alpha}$).

Before turning to the consequences of this observation we note that since Theorem ?? assumes $Z_0^*(t) = V_0 e^{\lambda_0 t}$ and then derives a representation for $V_1 = \lim_{t \rightarrow \infty} e^{-\lambda_1 t} Z_1^*(t)$, it follows from Theorem ?? and induction that V_k is the sum of points in a Poisson process with mean measure $\mu(x, \infty) = C_k x^{-\alpha}$ where $\alpha = \lambda_{k-1}/\lambda_k$. Fortunately the value of the constant is not important for the measures of diversity we consider.

10.1 Simpson's index

We define Simpson's index to be the probability two randomly chosen individuals in wave k are descended from the same mutation. In symbols,

$$R = \sum_{i=1}^{\infty} \frac{X_i^2}{V_k^2}$$

where $X_1 > X_2 > \dots$ are points in the Poisson process and V_k is the sum. The result for the mean is much simpler than one could reasonably expect.

Theorem 11. *$ER = 1 - \alpha$ where $\alpha = \lambda_{k-1}/\lambda_k$ for wave k .*

To prove this we apply results of Fuchs, Joffe and Teugels (2001) who considered

$$R_n = \sum_{i=1}^n \frac{Y_i^2}{S_n^2}$$

where the Y_i and S_n are as in Theorem ??, and showed that

$$\lim_{n \rightarrow \infty} ER_n = 1 - \alpha$$

To complete the proof one has to show that $\lim_{n \rightarrow \infty} ER_n = ER$.

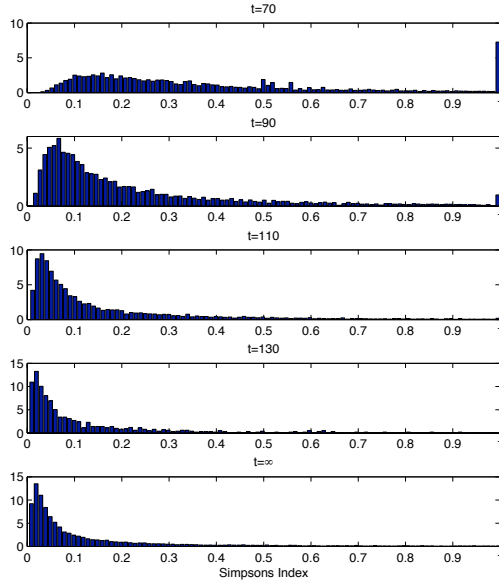


Figure 2: Empirical distribution of Simpson's Index for wave 1 at times $t = 70, 90, 110, 130, \infty$ when $1 - \alpha = 1/11$.

Our next topic is the distribution of R . Figure ?? shows a histogram of Simpson's index for wave 1 in a number of simulations. Logan, Mallows, Rice and Shepp (1973) considered the "self-normalized sums"

$$S_n(p) = \frac{\sum_{i=1}^n X_i}{(\sum_{j=1}^n X_j^p)^{1/p}}$$

which has $S_n(2) = R_n^{-1/2}$. They proved convergence in distribution and identified the Fourier transform of the limit. Despite considerable effort they were not able to calculate the limiting density f of $S_n(2)$, but in the case of interest to us they could infer that (see their (5.7), (5.9), and the caption of Figure 2)

$$\begin{aligned} f(y) &\sim ae^{-by^2} \quad \text{as } y \rightarrow \infty \\ &\sim \pi^{-1} \sin(\pi\alpha)(y-1)^{1-\alpha} \quad \text{as } y \downarrow 1 \end{aligned}$$

Differentiating $P(R_n < x) = P(S_n(2) > x^{-1/2})$ we see that the density g of R has

$$g(x) = \frac{1}{2} x^{-3/2} f(x^{-1/2})$$

and hence we have

$$\begin{aligned} g(x) &\sim \frac{a}{2} x^{-3/2} \exp(-b/x) \quad \text{as } x \downarrow 0 \\ &\sim c(1-x)^{1-\alpha} \quad \text{as } x \uparrow 1 \end{aligned}$$

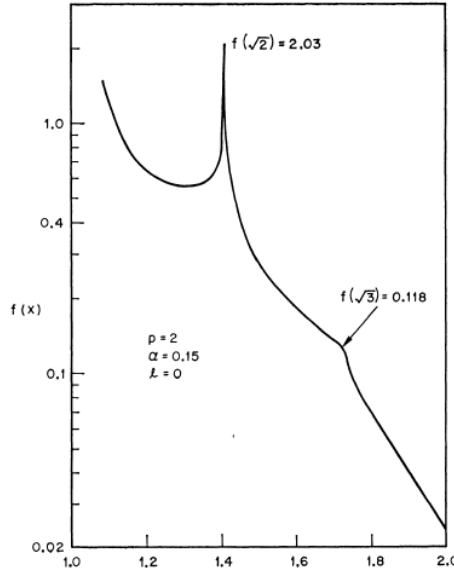


Figure 3: Density function for $S_n(2)$ when $\alpha = 0.15$

For the second formula note that $x^{-1/2}$ has derivative $-1/2$ at 1 so $x^{-1/2}-1 \sim (1-x)/2$ as $x \uparrow 1$.

Figure gives a picture of the density f in a special case. The constant ℓ which gives the relative size of the left tail of the distribution is 0, i.e., our situation with nonnegative random variables. The bumps at $\sqrt{2}$ and $\sqrt{3}$ suggests there will not be a simple formula for the density function

10.2 Largest clone

Using the notation of Theorem ?? let $U_n = \max_{1 \leq i \leq n} Y_i/S_n$ be the contribution of the largest term to the sum. Continuing to travel back in time, Darling (1952) has shown the following (see his Theorem 5.1)

Theorem 12. *As $n \rightarrow \infty$, $U_n^{-1} \rightarrow T$ where T has characteristic function $e^{it}/f_\alpha(t)$ where*

$$f_\alpha(t) = 1 + \alpha \int_0^1 (1 - e^{itu})u^{-(\alpha+1)} du$$

One cannot invert the characteristic function, but one can compute the moments

$$ET = 1/(1 - \alpha) \quad \text{and} \quad \text{var}(T) = 2/(1 - \alpha)^2(2 - \alpha)$$

It is remarkable that ET is so simple. Unfortunately $T = \lim U_n^{-1}$. To help interpret the next graph note that since $1/t$ is convex, Jensen's inequality implies $E(1/T) > 1/ET = 1 - \alpha$.

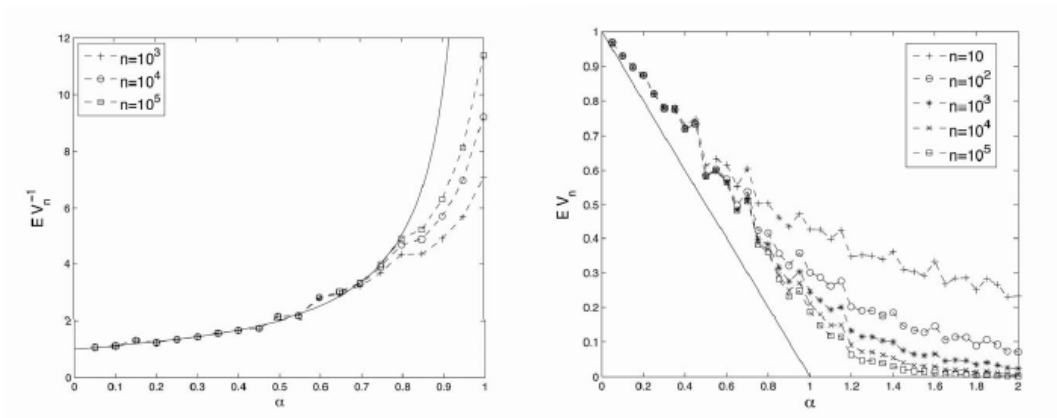


Figure 4: Monte Carlo estimates for $E(1/U_n)$ and EU_n plotted versus $1/(1 - \alpha)$ and $1 - \alpha$.

References

- Bozic I., Antal T. , Ohtsuki H., Carter H., Kim D., et al. (2009) Accumulation of driver and passenger mutations during tumor progression.
- Darling, D.A. (1952) The role of the maximum term in the sum of independent random variables. *Trans. AMS* 73, 95–107
- Durrett, R., Foo, J., Leder, K., Mayberry, J., Michor, F. (2010) Evolutionary dynamics of tumor progression with random fitness values. *Theor. Popul. Biol.* 78, 54–66
- Durrett, R., Foo, J., Leder, K., Mayberry, J., and Michor, F. (2011) Intratumor heterogeneity in evolutionary models of tumor progression. *Genetics*, to appear
- Durrett, R., and Moseley, S. (2010) Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Pop. Biol.* 77, 42-48
- Fuches, A., Joffee, A., and Teugels, J. (2001) Expectation of the ratio of the sum of squares to the square of the sum: exact and asymptotic results. *Theory Probab. Appl.* 46, 243–255
- Haeno, H., Iwasa, Y., and Michor, F. (2007) The evolution of two mutations during clonal expansion. *Genetics*. 177, 2209–2221
- Iwasa, Y., Nowak, M.A., and Michor, F. (2006) Evolution of resistance during clonal expansion. *Genetics*. 172, 2557–2566
- Leder, K., Foo, J., Skaggs, B., Gorre, M., Sawyers, C.L., and Michor, F. (2011) Modelling diversity in chronic myeloid leukemia and its therapeutic implications.
- Logan, B.F., Mallows, C.L., Rice, S.O., Shepp, L.A. (1973) Limit distributions of self-normalized sums. *Ann. Probab.* 1, 788–809